

STIC-Biotech/ChemLib

64186

From: Hunt, Jennifer
Sent: Wednesday, April 10, 2002 10:12 AM
To: STIC-Biotech/ChemLib
Subject: Seq Search for 09/234,208

Please search and interference search SEQ ID NO:1 and 2 of 09/234,208.

Thanks,

Jennifer Hunt
Patent Examiner, Art Unit 1642
CM1-8D06 (mailbox 8E12)
(703)308-7548

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

RECEIVED
APR 10 2002
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 4/11
Date Completed: 4/12
Searcher Prep/Review: 10
Clerical: _____
Online time: 10

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 2
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: es
WWW/Internet: _____
Other (specify): _____

1. 100-1000000
2. 100-1000000
3. 100-1000000
4. 100-1000000
5. 100-1000000

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:28:18 ; Search time 40.19 Seconds
(without alignments)
145.603 Million cell updates/sec

Title: US-09-234-208B-1

Perfect score: 418
Sequence: 1 GTHSLRPAPVAVPVLRLMQP.....VGRGPDPAHVAVNLSREG 79

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.Geneseq_1101:*

- 1: /SID8/gcgdata/geneseq/AA1980.DAT:*
- 2: /SID8/gcgdata/geneseq/AA1981.DAT:*
- 3: /SID8/gcgdata/geneseq/AA1982.DAT:*
- 4: /SID8/gcgdata/geneseq/AA1983.DAT:*
- 5: /SID8/gcgdata/geneseq/AA1984.DAT:*
- 6: /SID8/gcgdata/geneseq/AA1985.DAT:*
- 7: /SID8/gcgdata/geneseq/AA1986.DAT:*
- 8: /SID8/gcgdata/geneseq/AA1987.DAT:*
- 9: /SID8/gcgdata/geneseq/AA1988.DAT:*
- 10: /SID8/gcgdata/geneseq/AA1989.DAT:*
- 11: /SID8/gcgdata/geneseq/AA1990.DAT:*
- 12: /SID8/gcgdata/geneseq/AA1991.DAT:*
- 13: /SID8/gcgdata/geneseq/AA1992.DAT:*
- 14: /SID8/gcgdata/geneseq/AA1993.DAT:*
- 15: /SID8/gcgdata/geneseq/AA1994.DAT:*
- 16: /SID8/gcgdata/geneseq/AA1995.DAT:*
- 17: /SID8/gcgdata/geneseq/AA1996.DAT:*
- 18: /SID8/gcgdata/geneseq/AA1997.DAT:*
- 19: /SID8/gcgdata/geneseq/AA1998.DAT:*
- 20: /SID8/gcgdata/geneseq/AA2000.DAT:*
- 21: /SID8/gcgdata/geneseq/AA2001.DAT:*
- 22: /SID8/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406	97.1	79	AAV97241	HER-2 C-terminal e
2	355	84.9	79	AAV97239	HER-2 C-terminal p
3	351	84.0	420	AAV97240	truncated HER-2, p
4	86	20.6	459	AAV31404	HEV US and CMP-KDO
5	86	20.6	459	AAV31405	HEV US recombinant
6	83.5	20.0	888	AAW19857	Human semaphorin 2
7	83.5	20.0	888	AAW19916	A novel polypeptid
8	83.5	20.0	888	AAU12443	Human PRO4335 poly
9	82	19.6	132	AAV31402	HEV US and CMP-KDO
10	82	19.6	132	AAV31403	HEV US recombinant
11	81.5	19.5	1257	AAW46627	Neurocan core prot

12	78	18.7	122	AAV31386	HEV-US2 ORF3 prote
13	78	18.7	763	AAW31852	Mycoacterium tube
14	77	18.4	122	AAW93407	Swine HEV ORF 3 pr
15	77	18.4	329	AAW93933	Human protein sequ
16	75.5	18.1	198	AAW44557	Hepatitis C virus
17	75	17.9	530	AAW44300	Human ORFX ORF3064
18	75	17.9	1081	AAW95514	Human protein sequ
19	74.5	17.8	1246	AAW06296	Human transcriptio
20	74	17.7	616	AAW56941	Human prostate can
21	73.5	17.6	887	AAW19856	Rat semaphorin Z.
22	73	17.5	90	AAW54723	Arabidopsis thalia
23	73	17.5	148	AAW08181	Arabidopsis thalia
24	73	17.5	152	AAW54970	Arabidopsis thalia
25	73	17.5	164	AAW08180	Arabidopsis thalia
26	73	17.5	197	AAW08179	Arabidopsis thalia
27	73	17.5	639	AAW94220	Human protein sequ
28	73	17.5	1502	AAW39273	Human polypeptide
29	73	17.5	1565	AAW41059	Human polypeptide
30	72.5	17.3	833	AAW41835	Human ORFX ORF1599
31	72	17.2	122	AAV31383	HEV-US1 ORF3 prote
32	71.5	17.1	548	AAW07700	Human ETS2 repress
33	71.5	17.1	604	AAW06463	Derived protein of
34	71.5	17.1	604	AAW49848	Human pancreatic i
35	71.5	17.1	604	AAV33354	Human islet cell a
36	71.5	17.1	604	AAW80480	Islet cell antiod
37	71.5	17.1	604	AAV49322	Pancreatic islet c
38	71.5	17.1	615	AAW39081	Human polypeptide
39	71.5	17.1	2441	AAW79054	CREB binding prote
40	71.5	17.1	2441	AAW40058	Cellular transcrip
41	71.5	17.1	2441	AAV94252	Mouse nuclear CREB
42	71	17.0	148	AAW52235	Arabidopsis thalia
43	71	17.0	164	AAW52234	Arabidopsis thalia
44	71	17.0	197	AAW52233	Arabidopsis thalia
45	70.5	16.9	1026	AAW94315	Human protein sequ

ALIGNMENTS

RESULT 1
ID AAY97241 standard; Protein; 79 AA.
XX AAY97241;
AC AAY97241;
XX 04-DEC-2000 (first entry)
DT
XX
XX HER-2 C-terminal extracellular domain IIIA.
DE
XX
XX HER-2: erbB-2; oncogene; receptor-like tyrosine kinase; insertion;
KW extracellular domain IIIA; antagonist; intron 8; C-terminal extension;
KW truncated HER-2; p68; dimerization inhibitor; cytosolic.
XX
XX Homo sapiens.
OS
XX
XX WO200044403-A1.
PD
XX
XX 03-AUG-2000.
PF
XX
XX 20-JAN-2000; 2000WO-US01484.
PR
XX
XX 20-JAN-1999; 99US-0234208.
XX
XX (UYOR-) UNITV OREGON HEALTH SCI.
XX Doherty JK, Clinton GM, Adelman JP;
XX WPI: 2000-499287/44.
XX N-PSDB: AAA53783.
XX
XX Using polypeptides and antibodies that bind to the extracellular domain
XX of the receptor-like tyrosine kinase HER-2 to treat solid tumors of the
XX breast, lung, ovaries and colon

XX Example 11; Page 42-43; 46pp; English.

PS HER-2/neu (erbB-2) oncogene encodes a receptor-like tyrosine kinase. The

XX extracellular domain of p185-HER-2 is proteolytically shed from breast

CC carcinoma cells in culture and is found in serum of some cancer patients

CC and may be a serum marker of metastatic breast cancer. An alternative

CC HER-2 mRNA of 4.8 kb with a 274 bp insert (intron 8) has been

CC identified. The retained intron is in-frame and encodes a 79 amino acid

CC extension designated ECDIIIA (the present sequence), which is inserted at

CC residue 340 of p185-HER-2. The alternative mRNA predicts a truncated

CC HER-2 protein (approximately 68 kDa) that lacks the transmembrane and

CC intracellular domains (see AAY97240). p68HER-2 specifically binds to

CC p185-HER-2 without activating HER-2. It could therefore block

CC dimerization of p185-HER-2. The p68HER-2 polypeptide binds to a site on

CC the ECD of HER-2 that is different from the site of binding for

CC Herceptin (RTM) (a marketed humanized monoclonal antibody that is used

CC for the treatment of cancer and binds to the ECD of HER-2). The methods,

CC compositions, polypeptides and antibodies are used to treat solid

CC tumours such as breast cancer, small cell lung carcinoma, ovarian cancer

CC and/or colon cancer, especially where over-expression of HER-2 is

CC indicated.

XX Sequence 79 AA;

SQ

Query Match 97.1%; Score 406; DB 21; Length 79;

Best Local Similarity 97.5%; Pred. No. 2,1e-37;

Matches 77; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTHTSLPRRAAVVPLRMQGPRAHVPVLSFLRPSMDVLSAFYSIPLAPLSPTSPVSPVSV 60

Db 1 gthslprraavvplrmqgpahpvlslfrpswdvlsafysiplaplsptspvsvsv 60

QY 61 GRGPDPAHVAANVLSRYEG 79

Db 61 grgpdpahvavdlsryeg 79

RESULT 2

AA97239

ID AAY97239 standard; protein; 79 AA.

XX

AC AAY97239;

XX

DT 04-DEC-2000 (first entry)

XX

DE HER-2 C-terminal polymorphic extracellular domain IIIA.

XX

KW HER-2; erbB-2; oncogene; receptor-like tyrosine kinase; insertion;

KM extracellular domain IIIA; antagonist; intron 8; C-terminal extension;

KW truncated HER-2; p68; dimerization inhibitor; cytostatic.

XX

OS Homo sapiens.

XX

XX

XX Key Location/Qualifiers

FT Misc-difference 2 /note= "Preferably Ser"

FT Misc-difference 5 /note= "Preferably Pro"

FT Misc-difference 6 /note= "Preferably Leu"

FT Misc-difference 16 /note= "Preferably Gln"

FT Misc-difference 18 /note= "Preferably Leu"

FT Misc-difference 21 /note= "Changes from glycine"

FT Misc-difference 36 /note= "Preferably Ile"

FT Misc-difference 54 /note= "Preferably Arg"

FT Misc-difference 64

FT /note= "Preferably Leu"

FT Misc-difference 73 /note= "Preferably Asn"

XX

XX WO200044403-A1.

XX

XX 03-AUG-2000.

XX

XX 20-JAN-2000; 2000MO-US01484.

XX

XX 20-JAN-1999; 99US-0234208.

XX

XX (UYOR-) UNIV OREGON HEALTH SCI.

XX

XX Doherty JK, Clinton GM, Adelman JP;

XX

XX WPI; 2000-499287/44.

DR

XX

XX Using polypeptides and antibodies that bind to the extracellular domain

PT of the receptor-like tyrosine kinase HER-2 to treat solid tumors of the

PT breast, lung, ovaries and colon

XX

PS Claim 1; Page 39; 46pp; English.

XX

XX HER-2/neu (erbB-2) oncogene encodes a receptor-like tyrosine kinase. The

CC extracellular domain of p185-HER-2 is proteolytically shed from breast

CC carcinoma cells in culture and is found in serum of some cancer patients

CC and may be a serum marker of metastatic breast cancer. An alternative

CC HER-2 mRNA of 4.8 kb with a 274 bp insert (intron 8) has been

CC identified. The retained intron is in-frame and encodes a 79 amino acid

CC extension designated ECDIIIA (the present sequence), which is inserted at

CC residue 340 of p185-HER-2. The alternative mRNA predicts a truncated

CC HER-2 protein (approximately 68 kDa) that lacks the transmembrane and

CC intracellular domains (see AAY97240). p68HER-2 specifically binds to

CC p185-HER-2 without activating HER-2. It could therefore block

CC dimerization of p185-HER-2. The p68HER-2 polypeptide binds to a site on

CC the ECD of HER-2 that is different from the site of binding for

CC Herceptin (RTM) (a marketed humanized monoclonal antibody that is used

CC for the treatment of cancer and binds to the ECD of HER-2). The methods,

CC compositions, polypeptides and antibodies are used to treat solid

CC tumours such as breast cancer, small cell lung carcinoma, ovarian cancer

CC and/or colon cancer, especially where over-expression of HER-2 is

CC indicated.

XX

SQ Sequence 79 AA;

Query Match 84.9%; Score 355; DB 21; Length 79;

Best Local Similarity 87.3%; Pred. No. 8,2e-32;

Matches 69; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GTHTSLPRRAAVVPLRMQGPRAHVPVLSFLRPSMDVLSAFYSIPLAPLSPTSPVSPVSV 60

Db 1 gthslprraavvplrmqgpahpvlslfrpswdvlsafysiplaplsptspvsvsv 60

QY 61 GRGPDPAHVAANVLSRYEG 79

Db 61 grgpdpahvavdlsryeg 79

RESULT 3

AA97240

ID AAY97240 standard; protein; 420 AA.

XX

AC AAY97240;

XX

DT 04-DEC-2000 (first entry)

XX

DE Truncated HER-2, p68-HER-2.

XX

KW HER-2; erbB-2; oncogene; receptor-like tyrosine kinase; insertion;

KM extracellular domain IIIA; antagonist; intron 8; C-terminal extension;

KW truncated HER-2; p68; dimerization inhibitor; cytostatic.

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 124 /note= "represented as Arg"
XX FT Misc-difference 125 /note= "represented as Arg"
XX FT Misc-difference 343 /note= "represented as Arg"
XX FT Misc-difference 346 /note= "Preferably Ser"
XX FT Misc-difference 347 /note= "Preferably Pro"
XX FT Misc-difference 357 /note= "Preferably Leu"
XX FT Misc-difference 359 /note= "Preferably Gln"
XX FT Misc-difference 362 /note= "Preferably Leu"
XX FT Misc-difference 377 /note= "changes from glycine"
XX FT Misc-difference 395 /note= "Preferably Ile"
XX FT Misc-difference 405 /note= "Preferably Arg"
XX FT Misc-difference 414 /note= "Preferably Leu"
XX FT Misc-difference 414 /note= "Preferably Asn"
XX PN WO200044403-A1.
XX PD 03-AUG-2000.
XX PF 20-JAN-2000; 2000WO-US01484.
XX PR 20-JAN-1999; 99US-0234208.
XX PA (UYOR-) UNIV OREGON HEALTH SCI.
XX PI Doherty JK, Clinton GM, Adelman JP;
XX WP1; 2000-499287/44.
XX PT using polypeptides and antibodies that bind to the extracellular domain
XX PT of the receptor-like tyrosine kinase HER-2 to treat solid tumors of the
XX PT breast, lung, ovaries and colon
XX PS Claim 8: Page 39-40; 46pp; English.
XX CC HER-2/neu (erbB-2) oncogene encodes a receptor-like tyrosine kinase. The
XX CC extracellular domain of p185-HER-2 is proteolytically shed from breast
XX CC carcinoma cells in culture and is found in serum of some cancer patients
XX CC and may be a serum marker of metastatic breast cancer. An alternative
XX CC HER-2 mRNA of 4.8 kb with a 274 bp insert (intron 8) has been
XX CC identified. The retained intron is in-frame and encodes a 79 amino acid
XX CC extension designated ECDIITa (the present sequence), which is inserted at
XX CC residue 340 of p185-HER-2. The alternative mRNA predicts a truncated
XX CC HER-2 protein (approximately 68 kDa) that lacks the transmembrane and
XX CC intracellular domains (see AAY97240). p68HER-2 specifically binds to
XX CC p185-HER-2 without activating HER-2. It could therefore block
XX CC dimerization of p185-HER-2. The p68HER-2 polypeptide binds to a site on
XX CC the ECD of HER-2 that is different from the site of binding for
XX CC Herepentin (RTM) (a marketed humanized monoclonal antibody that is used
XX CC for the treatment of cancer and binds to the ECD of HER-2). The methods,
XX CC compositions, polypeptides and antibodies are used to treat solid
XX CC tumors such as breast cancer, small cell lung carcinoma, ovarian cancer
XX CC and/or colon cancer, especially where over-expression of HER-2 is
XX CC indicated.
XX SQ Sequence 420 AA:

```

Query Match

84.0%; Score 351; DB 21; Length 420;

```

Best Local Similarity 86.1%; Pred. No. 1.5e-30;
Matches 68; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
OY 1 GTHSLPRPAAVPVLRMQPGPAHPVLSFLRSPWDLVSAFYSILPLAPLSPTSVSPSV 60
DB 342 gthsxprpaavpvrpxpaphvlsflrpswdxvsafysilplaplptsvxspav 401
OY 61 GRGPPDAHVAVNLSRYEG 79
DB 402 grgppdahvavxlsryeg 420
RESULT 4
AA31404
ID AA31404 standard; Protein: 459 AA.
XX AA31404;
AC AA31404;
XX 12-OCT-1999 (first entry)
DE HEV US and CMP-KDO synthetase (CKS) fusion protein CKSORE32M-3.
XX Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
KW vaccine; passive immunisation.
XX Hepatitis E virus.
XX OS Hepatitis E virus.
XX PN WO9919732-A1.
XX PD 22-APR-1999.
XX PF 15-OCT-1998; 98WO-US21941.
XX PR 15-OCT-1997; 97US-0061199.
XX PA (ABBO) ABBOTT LAB.
XX PI Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
XX WP1; 1999-288017/24.
XX DR N-PSDB; AA200280.
XX PT Detection of United States isolates of hepatitis E virus
XX PS Example 10: Page 241-243; 260pp; English.
XX CC The invention provides a method for detecting a US (sub)type hepatitis E
XX CC virus (US-HEV), or its naturally occurring variants in a sample by
XX CC treatment with a binding partner specific for a marker of the virus, and
XX CC then detecting any complex formed. The method is used to diagnose
XX CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
XX CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
XX CC are also useful in vaccines or for passive immunisation. The polypeptides are
XX CC also used to raise specific antibodies (useful as immunoassay reagents).
XX CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
XX CC usual hybridisation and amplification assays for detecting infection.
XX SQ Sequence 459 AA:

```

```

Query Match 20.6%; Score 86; DB 20; Length 459;
Best Local Similarity 34.2%; Pred. No. 0.19;
Matches 27; Conservative 12; Mismatches 28; Indels 12; Gaps 4;
OY 1 GTHSLPRPAAVPVLRMQPGPAHPVLSFLRSPWDLVSAFYSILPLAPLSPTSVSPSV 57
DB 57 gthslprpaavpvrpxpaphvlsflrpswdxvsafysilplaplptsvxspav 113
OY 58 V-----SVGRPPDAHV 70
DB 114 vvdldpqlgrtgadgtael 132

```


[illegible]

PD		28-SEP-2000.	
XX	FE	01-MAR-2000; 2000WO-USO5601.	
XX	PR	23-MAR-1999; 99US-0125774.	
XX	PR	23-MAR-1999; 99US-0125778.	
XX	PR	24-MAR-1999; 99US-0125826.	
XX	PR	31-MAR-1999; 99US-0127035.	
XX	PR	05-APR-1999; 99US-0127706.	
XX	PR	21-APR-1999; 99US-0130359.	
XX	PR	27-APR-1999; 99US-0131272.	
XX	PR	27-APR-1999; 99US-0131291.	
XX	PR	04-MAY-1999; 99US-0132371.	
XX	PR	04-MAY-1999; 99US-0132379.	
XX	PR	04-MAY-1999; 99US-0132383.	
XX	PR	25-MAY-1999; 99US-0135750.	
XX	PR	08-JUN-1999; 99US-0138166.	
XX	PR	20-JUL-1999; 99US-0144791.	
XX	PR	03-AUG-1999; 99US-0146970.	
XX	PR	09-DEC-1999; 99US-0170262.	
XX	PA	(GETH) GENENTECH INC.	
XX	PI	Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;	
XX	PI	Stewart T, Watanabe CK, Wood WI, Zhang Z;	
DR	DR	WPI: 2000-628263/60.	
XX	N-PSDB:	AAA96343.	
XX	PT	Novel secreted and transmembrane polypeptides useful for diagnosing	
PT	PT	tumour in a mammal, for identifying agonists and antagonists of the	
PT	PT	polypeptide and for therapeutic use	
XX	XX	Claim 12; Fig 16; 222pp: English.	
XX	CC	The present sequence represents a secreted or transmembrane polypeptide.	
CC	CC	The specification describes polypeptides designated PRO1484, PRO4334,	
CC	CC	PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4405,	
CC	CC	PRO4356, PRO4380, PRO4380, PRO4354, PRO4408, PRO5337, PRO4425, PRO5990,	
CC	CC	PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is	
CC	CC	useful for diagnosing tumour in a mammal. The polypeptides, their	
CC	CC	agonists and antagonists are useful treating a condition associated with	
CC	CC	expression or activity of the polypeptide. Conditions treated include	
CC	CC	obesity, diabetes or hyper-or hypo-insulinemia. The polypeptides are	
CC	CC	capable of inducing proliferation of mammalian kidney mesangial cells	
CC	CC	and are therefore useful for treating kidney disorders associated with	
CC	CC	decreased mesangial cell function such as Bergrers disease or other	
CC	CC	nephropathies associated with Schonlein-Henoch purpura, celiac disease,	
CC	CC	dermatitis herpetiformis or Crohns disease. The nucleic acids may be used	
CC	CC	to generate transgenic animals for use in development and screening of	
CC	CC	therapeutically useful reagents and also for chromosome identification	
CC	CC	and tissue typing.	
SO	Sequence	888 AA;	
OY	Query Match	20.0%; Score 83.5; DB 21; Length 888;	
Db	Best Local Similarity	35.4%; Pred. No. 0.76; Mismatches 33; Indels 11; Gaps 5;	
OY	Matches	26; Conservative 7; Mismatches 33; Indels 11; Gaps 5;	
OY	1 GTH----	SLFRPAAVPYELRMQGPAAHVLSLRP-SWD---LVSAFYSLPLAAPSPT 51	
Db	698 gphaldsgllptpcqtgtpipqkrlptc-hpnphaigprawdhgnhlppasasslllila	756	
OY	52 SVTSPVSVSGKPPDDAHV 70		
Db	757 rapeqppeape-pcpdgri 774		
RESULT	8		
ID	AAU12443	standard: Proteiin: 888 AA	
ID	AAU12443		

XX AAU12443;
 AC XX 24-OCT-2001 (first entry)
 XX
 DE Human PRO4353 polypeptide sequence.
 XX
 KW Human secretory and transmembrane; PRO: mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200140466-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 01-DEC-2000; 2000WO-US32678.
 XX
 PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 09-DEC-1999; 99US-0170262.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30939.
 PR 30-DEC-1999; 99WO-US31243.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.
 XX
 XX (GETH) GENENTECH INC.
 PA
 PI Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerlitsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2001-408281/43.
 DR N-PSDB; AAS21515.
 XX
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX
 PS Claim 12; Fig 544; 813pp; English.
 XX
 CC AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the

CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIa. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 888 AA:
 XX
 QY 1 GTH----SLPPRAAVPVLRMQPGPAHPVLSFLRP-SMD---LVSAPYSPLAPLSPT 51
 Db 698 gphldsglllppeqtpipqkrlpvp-nphphalgrawdhpllpasasslllllpa 756
 QY 52 SVPISEFVSVGRGPPDPDAHV 70
 Db 757 rapqppapqge-plpddgrl 774
 XX
 RESULT 9
 AAY31402
 ID AAY31402 standard; Protein; 132 AA.
 XX
 AC AAY31402;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE HEV US and CMP-KDO synthetase (CKS) fusion protein.
 XX
 KW Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
 KW vaccine; passive immunisation.
 XX
 OS Hepatitis E virus.
 OS
 PN WO9919732-A1.
 XX
 PD 22-APR-1999.
 XX
 PF 15-OCT-1998; 98WO-US21941.
 XX
 PR 15-OCT-1997; 97US-0061199.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
 XX
 DR WPI; 1999-288017/24.
 DR N-PSDB; AAZ00278.
 XX
 PT Detection of United States isolates of hepatitis E virus
 PT
 PS Example 10; Page 240; 260pp; English.
 XX
 CC The invention provides a method for detecting a US (sub)type hepatitis E
 CC virus (US-HEV), or its naturally occurring variants in a sample by
 CC treatment with a binding partner specific for a marker of the virus, and
 CC then detecting any complex formed. The method is used to diagnose
 CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
 CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
 CC are useful in vaccines or for passive immunisation. The polypeptides are
 CC also used to raise specific antibodies (useful as immunoassay reagents).
 CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
 CC usual hybridisation and amplification assays for detecting infection.
 XX


```

FT Modified-site 857 /note= "O-linked glycosylation site"
FT Modified-site 859 /note= "O-linked glycosylation site"
FT Modified-site 861 /note= "O-linked glycosylation site"
FT Modified-site 863 /note= "O-linked glycosylation site"
FT Modified-site 905 /note= "O-linked glycosylation site"
FT Modified-site 913 /note= "O-linked glycosylation site"
FT Modified-site 933 /note= "O-linked glycosylation site"
FT Peptide 942 /note= "used for primer design"
FT Binding-site 944 /note= "Chondroitin sulphate attachment site"
FT Modified-site 950 /note= "O-linked glycosylation site"
FT Modified-site 967 /note= "N-linked glycosylation site"
FT Modified-site 1136 /note= "N-linked glycosylation site"
FT Peptide 1147 /note= "used in primer design"
FT Modified-site 1164 /note= "N-linked glycosylation site"
FT Modified-site 1164 /note= "N-linked glycosylation site"
PN WO9403601-A.
PD 17-FEB-1994.
XX
XX
XX 03-AUG-1993; 93WO-US07306.
XX
XX 03-AUG-1992; 92US-0922911.
XX
XX (UYNV ) UNIV NEW YORK STATE.
XX
XX Margolis RK, Margolis RU, Rauch U;
PI
XX
XX WPI: 1994-065690/08.
DR N-PSDB; AA057710.
XX
XX Eukaryotic neurocan polypeptide(s) with epidermal growth factor,
PT lectin or complement binding activity - used in the diagnosis,
PT treatment or research of hypersensitivity and allergic diseases
XX
XX Claim 1; Page 69-75; 105pp; English.
XX
XX This sequence represents a neurocan polypeptide. This protein has
CC several biological activities, including cell adhesion, leukocyte-
CC endothelial cell recognition, tissue-related inflammation allergies,
CC cellular and/or humoral hypersensitivity, trauma, neuronal
CC development, and cell transport and/or infection. Compositions
CC containing them can be used as modulators of these conditions, and
CC may be used as therapeutic, diagnostic, and/or research tools.
CC Neurocan peptides can be used to mimic proteins, such as lectins,
CC cell adhesion molecules, versicans, aggrecans or gelsolins, as
CC receptor or effector subtypes. The protein can be used to treat
CC diseases involving a qualitative or quantitative pathological
CC abnormality of cell adhesion or leukocyte-endothelial cell recognition,
CC or a functionally associated molecule such as a membrane cytoplasmic
CC protein, lipid, carbohydrate, saccharide, nucleoside, enzyme or ion.
XX
XX Sequence 1257 AA:
SQ

```

Query Match 19.5%; Score 81.5; DB 15; Length 1257;
Best Local Similarity 35.0%; Pred. No. 1.9;
Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

```

QY 4 SLLPRRAAVPVLPMQPG---FAHPVLSFLR-----PSMDIVSAFYSLPLAPLS--PT 51
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 610 ssipsealseavsqspdgspdfivamlrpk1w1p1hstlvnpvslp1spasplps 669
QY 52 SVP-----ISPVSVGRGPPDP 67

```

```

DB 670 svpeegavrvpsfg-aedpe 688
| | | : | | | : | : |
RESULT 12
AAV31386
ID AAV31386 standard; Protein; 122 AA.
XX
XX AAV31386;
AC
XX
XX 12-OCT-1999 (first entry)
DE
XX HEV-US2 ORF3 protein.
XX
XX Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
KW vaccine; passive immunisation.
XX
XX Hepatitis E virus.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 97 /label= unknown
FT /note= "encoded by GYT"
FT
FT
PN WO9919732-A1.
PD 22-APR-1999.
XX
XX
XX 15-OCT-1998; 98WO-US21941.
XX
XX 15-OCT-1997; 97US-0061199.
XX
XX (ABBO ) ABBOTT LAB.
PA
XX Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GC;
PI
XX
XX WPI: 1999-288017/24.
DR N-PSDB; AA020267.
XX
XX Detection of United States isolates of hepatitis E virus
PT
XX
XX Claim 18; Page 223; 260pp; English.
XX
XX The invention provides a method for detecting a US (sub)type hepatitis E
CC virus (US-HEV), or its naturally occurring variants in a sample by
CC treatment with a binding partner specific for a marker of the virus, and
CC then detecting any complex formed. The method is used to diagnose
CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC are useful in vaccines or for passive immunisation. The polypeptides are
CC also used to raise specific antibodies (useful as immunoassay reagents).
CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC usual hybridisation and amplification assays for detecting infection. The
CC present sequence represents a HEV-US2 ORF3 protein.
XX
XX Sequence 122 AA:
SQ

```

Query Match 18.7%; Score 78; DB 20; Length 122;
Best Local Similarity 36.1%; Pred. No. 0.31;
Matches 22; Conservative 10; Mismatches 23; Indels 6; Gaps 3;

```

QY 1 GTHSLPRRAAVPVLPMQPGFAHPVLSFLRPSMDIVSAFYSLPLAPL---SPTSPISP 57
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 55 gvtgllsps--psplfqltps--pmsfthngjlelaldspaplxp1gvtspasplpp 111
QY 58 V 58
DB 112 v 112

```

RESULT 13
AAM31852

```

ID AAW31852 standard; Protein; 763 AA.
XX
AC AAW31852;
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 74 kDa protein.
XX
KM Tuberculosis; mycobacteria; infection; diagnosis;
KM antimycobacterial; antituberc; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9741252-A2.
XX
PD 06-NOV-1997.
XX
PF 18-APR-1997; 97WO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PA (GBFB ) GBF GBS BIOTECH FORSCHUNG GMBH.
XX
PI Espitia C, Honisch C, Moreno C, Singh M;
XX
DR WPI; 1997-549750/50.
XX
DR N-PSDB; AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis -
PT used for diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents
XX
PS Claim 5; Fig 13; 55pp; English.
XX
CC 'This novel 74 kDa protein is encoded by an open reading frame of
CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
XX
SQ Sequence 763 AA;

Query Match 18.7%; Score 78; DB 18; Length 763;
Best Local Similarity 34.8%; Pred. No. 2.5;
Matches 24; Conservative 4; Mismatches 33; Indels 8; Gaps 2;

QY 4 SLPPPAAPVPLRMQPGAPHPVLSFLRPSMDLVSAFYSLLPLASP---TSVPISPV 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 9 alappppppppaepkspkfpfpappppcmvlvsaaappcpappppkpkapfpvp 68
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 VGRGPDPPDA 68
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 69 ----pappa 73
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
AAW93407
ID AAW93407 standard; Protein; 122 AA.
XX
AC AAW93407;
XX
DT 11-JUN-1999 (first entry)
XX
DE Swine HEV ORF 3 protein.
XX

```

```

KM Swine hepatitis E virus; HEV; cross-reaction; antibody; human; therapy;
KM vaccine; immunise; infection; detection; diagnosis; prevention.
XX
OS Hepatitis E virus.
XX
PN WO9904029-A2.
XX
PD 28-JAN-1999.
XX
PF 17-JUL-1998; 98WO-US14665.
XX
PR 18-JUL-1997; 97US-0053069.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Emerson SU, Meng X, Purcell RH;
XX
DR WPI; 1999-132270/11.
XX
PT New isolated swine hepatitis E virus - used to develop products for
PT the diagnosis, prevention and treatment of hepatitis E virus
PT infection in mammals, particularly humans
XX
PS Example 1; Fig 3B; 70pp; English.
XX
CC This invention describes a swine hepatitis E virus (HEV) and its natural
CC mutants which are capable of cross-reacting with antibodies reactive
CC with a human HEV strain or natural mutants. The HEV and the proteins
CC can be used in vaccines for immunising against HEV infection. The swine
CC HEV can be used in humans to prevent possible infection by human HEV. The
CC swine HEV can also be used as a therapeutic treatment for infection by
CC other strains of HEV. The swine HEV can also be used for the production
CC of antibodies which can be used in therapy, detection and diagnosis. The
CC products can also be used for determining the susceptibility of cells or
CC organs to infection with swine HEV. The swine HEV is particularly useful
CC for the development of agents for the prevention, treatment and detection
CC of human HEV because it is not a human virus and thus can be handled both
CC experimentally and clinically without fear of severe infection and/or
CC contamination.
XX
SQ Sequence 122 AA;

Query Match 18.4%; Score 77; DB 20; Length 122;
Best Local Similarity 36.1%; Pred. No. 0.4;
Matches 22; Conservative 10; Mismatches 23; Indels 6; Gaps 3;

QY 1 GTHSLRPAAVPLRMQPGAPHPVLSFLRPSMDLVSAFYSLLPLAPL--SPTSVPISP 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 55 gvtglllpsps-ppslfllqprslp-msfhnpglftalsraplaplgvtspsapplp 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 V 58
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 112 v 112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
AAB93933
ID AAB93933 standard; Protein; 329 AA.
XX
AC AAB93933;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13931.
XX
KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX

```

```

XX 28-JUL-2000; 2000EP-0116126.
PF
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13931; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 329 AA;

```

Query Match 18.4%; Score 77; DB 22; Length 329;
 Best Local Similarity 34.3%; Pred. No. 1.2;
 Matches 24; Conservative 8; Mismatches 24; Indels 14; Gaps 2;

```

QY 8 RPAAPVPLRMQGPAPAHVPLSLRPSWDLVSAFYSLLAPLSPTSVPSVSGRGDPD 67
   ||: | :||| | | ||| | | | | | | | | | | | | | | | | | |
Db 231 rpskydvrgiqkpgpak-----vppqslapapla--svpsapsapgppepp 276
QY 68 AHYAVNLSRY 77
   | : | | |
Db 277 asisfntpey 286

```

Search completed: April 11, 2002, 09:29:08
 Job time: 50 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:28:23 ; Search time 25.01 Seconds

(without alignments)
240.615 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418

Sequence: 1 GTRSLPRAAVPPLRMQP.....VGRGPDPAHVAVNLSRYEG 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.5	19.5	1257	2 S28764	neurocan precursor
2	81	19.4	200	2 H84715	probable phytoeyan
3	81	19.4	1006	2 G68292	hypothetical prote
4	80	19.1	1834	1 JDMU1	DNA-directed RNA p
5	80	19.1	1840	2 G85422	hypothetical prote
6	78	18.7	249	2 S72619	hypothetical prote
7	77.5	18.5	217	2 T51031	related to finger
8	77	18.4	503	2 T19319	hypothetical prote
9	76.5	18.3	1110	2 T19673	hypothetical prote
10	76	18.2	356	2 A96826	hypothetical prote
11	76	18.2	848	2 S48273	probable transcrip
12	75	17.9	189	2 D49600	genome-linked prot
13	75	17.9	1952	2 T48814	hypothetical prote
14	74.5	17.8	894	2 T13029	beta-adaptin homol
15	73	17.5	1520	2 T100273	hypothetical prote
16	72.5	17.3	518	2 F70831	probable PPE prote
17	72	17.2	440	2 A44081	kappa-type oploid
18	72	17.2	1914	2 T42635	tenascin Y precurs
19	71.5	17.1	2441	2 S39161	CREB-binding prote
20	71	17.0	906	2 A71438	probable resistanc
21	70.5	16.9	291	2 S27721	hypothetical prote
22	70.5	16.9	443	2 J27877	hypothetical prote
23	70.5	16.9	788	1 J2V1H8	DNA-directed DNA p
24	70.5	16.9	1268	1 S52781	neurocan - mouse
25	70	16.7	491	2 S41182	DNA-directed RNA p
26	70	16.7	650	2 S41181	DNA-directed RNA p
27	70	16.7	954	2 E86174	protein Flp19.26
28	70	16.7	2187	2 T30826	nascent polypeptid
29	70	16.7	2318	2 S45306	notch 3 protein -

30	70	16.7	2321	2 S78549	notch3 protein - h
31	69.5	16.6	433	2 T09284	TEA domain-contain
32	69.5	16.6	608	2 A46312	gag polyprotein -
33	69.5	16.6	684	2 T25603	hypothetical prote
34	69	16.5	148	2 T38881	caudal-type homeot
35	69	16.5	280	2 A75526	S-layer-like array
36	69	16.5	1121	2 T02764	myosin-I binding p
37	69	16.5	1353	2 T00249	hybrid proline-ric
38	68.5	16.4	401	2 T51407	arabinogalactan-pr
39	68.5	16.4	461	2 T10285	probable spermidin
40	68.5	16.4	554	2 T36345	probable potassum
41	68.5	16.4	1017	2 T31354	CD4 peptide synthe
42	68.5	16.4	7463	2 T36248	hybrid proline-ric
43	68	16.3	301	2 J01663	hypothetical prote
44	68	16.3	444	2 E83802	mucin JUL7 - human
45	68	16.3	543	2 S35047	

ALIGNMENTS

```

RESULT 1
S28764
neurocan precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C:Accession: S28764
R:Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
J. Biol. Chem. 267, 19536-19547, 1992
A:Title: Cloning and primary structure of neurocan, a developmentally regulated, aggr
A:Reference number: S28764; MUID:92406907
A:Accession: S28764
A:Molecule type: mRNA
A:Residues: 1-1257 <RNU>
A:Cross-references: EMBL:M97161; NID:9205649; PIDN:AC37679.1; PID:9205650
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein
F:1-22/Domain: signal sequence #status predicted <MAT>
F:23-1257/Product: neurocan #status predicted <MAT>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK2>
F:364-366/Region: cell attachment (R-G-D) motif
F:953-984/Domain: EGF homology <EGF>
F:1029-1149/Domain: C-type lectin homology <LCH>
F:1156-1212/Domain: complement factor H repeat homology <FRH>
F:1121,339,737,967,1164/Binding site: carboxyrate (Asn) (covalent) #status predicted
F:372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F:944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 19.5% Score 81.5; DB 2; Length 1257;
Best local Similarity 35.0%; Pred. No. 2.7;
Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

OY 4 SLPPRAAVPPLRMQP---PAHPVLSFLR-----PSMDVSAFYSPLAPLS--PT 51
DB 610 SSIPSEALSAVSLQASPDGSDPRIVAMLRKMLPHSLIVNVSPIPLSPAPLS 669
OY 52 SVP---ISPVSVGRGPPD 67
DB 670 SVPEQAVRPVSFG-AEDPE 688

RESULT 2
H84715
probable phytoeyanin [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84715
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Mofat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999

```

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*
A:Reference number: AB4420; MIMD:20083487
A:Accession: H84715
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1..200 <STO>
A:Cross-references: GB:AE002093; NID:g3746072; PIDN:AA6C3847.L; GSPDB:GM00139
C:Genetics:
A:Gene: At2g31050
A:Map position: 2

	Query Match	19.4%	Score 81:	DB 2:	Length 200;
Oy	Best Local Similarity	34.3%;	Pred. No.	0.37;	
Db	Matches	23; Conservative	5; Mismatches	23; Indels	16; Gaps 2
Oy	9 PAAVVPRLMOGPAHPVFLRPSWDLVSAFYSLEPLAPLSPTSPVISGVSGRPDPA	68			
			:	:	:
Db	132 PVAAPV-----PGVPKPPSSFSPPS-----QSPLAASPVNHAHVQIOMGPSPAP	175			
Oy	69 HVAVNLS	75			
	:				
Db	176 HSAASNLS	182			

RESULT 3
G86292
hypochemical protein AAF82153.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86292
R:Thellogists, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
A.A.; Liu, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzalli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: G86292
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1006 <SNO>
A:Cross-references: GB:AE005172; NID:g8927662; PIDN:AAF82153.1; GSPDB:GM00141
C:Genetics:
A:Map position: 1

```
Query Match Similarity      19.4%; Score 81; DB 2; Length 1006;
Best Local Similarity       31.4%; Pred No. 2.3;
Matches    22; Conservative   5; Mismatches    21; Indels    22; Gaps     2;
```

```
Oy      6 LRPAAVPLRLMQGPAPHPVLSPLRPSWDLVSAEYSLPFLASPTSVIPSPVSGRGPD 65
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      336 LEPATLPPPLDLPPLPPP-----SLVPFCSP---PPPIIYNCAKP 373
```

```
Oy      66 PDAAVAANLS 75
          |         | : |
Db      374 PCCVTGVGS 383
```

```
RESULT      4
JDM01
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Arabidopsis thaliana
N:Alternate names: DNA-directed RNA polymerase II 205K chain; protein FAbI.70
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 31-Mar-1993 #sequence_revision 11-Jun-1999 #text_change 20-Aug-1999
C:Accession: r04690; S12071; S27346; S11960
R:Beyan, M.; Rose, W.; Hempel, S.; Entian, K.D.; Hohnseil, J.; Mexes, H.W.; Mayer, K.F.J.
Submitted to the Protein Sequence Database, October 1998
```

A:Reference number: 215380
A:Accession: T04690
A:Molecule type: DNA
A:Residues: 1-1834 <BEV>
A:Cross-references: EMBL:AL031986
A:Experimental source: cultivar Columbia; BAC clone *FAB14*
R:Navrath, C.; Schell, J.; Koncz, C.
Mol. Gen. Genet. 223, 65-75, 1990
A>Title: Homologous domains of the largest subunit of eucaryotic RNA polymerase II are
A:Reference number: SI2071, MUID:91080867

A:Residues: 1121L,'S',442L-732L,'D',734L-1055L,'R',1057L-1714L,'SPSPSPY',1715L-1834L <NAN1>
A:Cross-references: EMBL:X52954; NID:916504; PIDN:CAA37130.1; PID:916505
A:Experimental source: cv, Columbia
A:Note: the authors translated the codon AGC for residue 1755 as Arg
A:Accession: S27346
A:Molecule type: mRNA
A:Residues: 510-732L,'D',734L-1055L,'R',1057L-1714L,'SPSPSPY',1715L-1834L <NAN2>
A:Experimental source: cv, Columbia
A:Riedtlich, M.A.; Prenger, J.P.; Gullfoyle, T.J.
Plant Mol. Biol. 15, 207-223, 1990
A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A
A:Reference number: S11960; MUID:9135869
A:Accession: S11960
A:Molecule type: DNA
A:Residues: 1116L,125L-192L,'NSKEE',198L-297L,'R',299L-302L,'R',304L-403L,'VDYGHPEPP
A:Cross-references: EMBL:X52494; NID:916493; PIDN:CAA36735.1; PID:916494
A:Note: the authors translated the codon CCT for residue 1083 as Ala
A:Genetics:
A:Gene: rpl1215, RPB1
A:Map position: 4
A:Introns: 28/3; 123/3; 218/3; 271/3; 325/3; 411/1; 440/3; 552/3; 648/2; 734/3; 1755/
A:Note: F4B14.70
C:Superfamily: human DNA-directed RNA polymerase II largest chain
C:Keywords: DNA binding; nucleotidyltransferase; phosphoprotein; tandem repeat; trans
E:66-109/Region: zinc finger CCCC motif
E:1531-1812/Region: 7-residue repeats

	Query Match	Similarity	Score	DB 1	Length	1834;
Best Local	Similarity	36.4%	Pred. No.	5.9;		
Matches	28;	Conservative	3;	Mismatches	36;	Indels 10; Gaps 3.
Qy	7	PPPAVAVPVLRLRQ-----GPAHVLSFLRPPSDVLVSAYSLDPLAPLSPVSPVGR	62			
Db	1730	PSIAIAPSNAKLSPASPSYPTSPNTSPISPSISPSIS-----PSSPIYSSPSPISSGA	1785			
Qy	63	GPD--PDAHVAVNLSRY	77			
Db	1786	SPDIYSPSACISPTLPGR	1802			

RESULT 5
G85422
hypothetical protein AT4g35800 [imported] - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
C.Accession: G85422
R.anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory
A.Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A.Reference number: A85001; MUID:20083488
A.Accession: G85422
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1840 <SP>
A.Cross-references: GB:NC_001266; NID:g7270532; PIDN:CAB81489.1; GSPDB:GN00140
C.Genetics:
A.Gene: AT4g35800
A.Map position: 4
C.Superfamily: human DNA-directed RNA polymerase II largest chain

Query Match 19.1%; Score 80; DB 2; Length 1840;
 Best Local Similarity 36.4%; Pred. No. 5.9;
 Matches 28; Conservative 3; Mismatches 36; Indels 10; Gaps 3;

OY 7 PPAAPVPLRMQ-----GPAHPVLSFLRPSMDLVSAFYSLAPLSPTSPVSPVSGR 62
 DB 1736 PSIAVSPSNAARLSPASPSPTSPNYSPTSPVSPVSGR-----PSSPTSPVSPVSGR 1791
 OY 63 GPD--PDAAVAVNLSRX 77
 DB 1792 SPDYSPSAGYSPVLPQY 1808

RESULT 6
 S72619
 Hypothetical protein 1a - anthracnose fungus (Colletotrichum gloeosporioides) retrotrans
 C:Species: Glomerella cingulata, Colletotrichum gloeosporioides
 C:Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 17-Mar-1999
 C:Accession: S72619
 R:He, C.; Nourse, J.P.; Kelenyi, S.; Irwin, J.A.G.; Manners, J.M.
 Mol. Gen. Genet. 252, 320-331, 1996
 A:Title: Cg1: a non-UTR retrotransposon with restricted distribution in the fungal phyt
 A:Reference number: S72619; MUID:96439839
 A:Accession: S72619
 A:Molecule type: DNA
 A:Residues: 1-249 <HEA>
 A:Cross-references: EMBL:L76172
 A:Experimental source: biotype B, isolate U062
 A:Note: In the authors' translation residues 1-10 are not shown
 C:Genetics:
 A:Mobile element: retrotransposon Cg1

Query Match 18.7%; Score 78; DB 2; Length 249;
 Best Local Similarity 34.7%; Pred. No. 0.94;
 Matches 25; Conservative 7; Mismatches 26; Indels 14; Gaps 3;

OY 5 ILPR-----PAAPVPLRMQGPAPVLSFLRPSM---DIVSAFYSLAPLSPTSPV 54
 DB 24 LVPRVCHPNTYFQPTFOKEPSAMASQPAAPSGCGIFGLSSMHNLPTPPPTSLP 83
 OY 55 ISPVSGRGPPD 66
 DB 84 PRP-----RGLTP 91

RESULT 7
 T51031
 related to finger protein xfg 68 [imported] - Neurospora crassa
 N:Alternate names: protein B15120.10
 C:Species: Neurospora crassa
 C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
 C:Accession: T51031
 R:Schulte, U.; Aign, V.; Hohnsels, J.; Brandt, P.; Farmmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, July 2000
 A:Reference number: 225286
 A:Accession: T51031
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-217 <SCH>
 A:Cross-references: EMBL:AL389900; GSPDB:GN00116; NCSP:B15120.10
 A:Experimental source: BAC clone B15120; strain OR74A
 C:Genetics:
 A:Gene: NCSP:B15120.10
 A:Map position: 6

Query Match 18.5%; Score 77.5; DB 2; Length 217;
 Best Local Similarity 37.5%; Pred. No. 0.9;
 Matches 21; Conservative 5; Mismatches 21; Indels 9; Gaps 2;

OY 7 PPAAPVPLRMQGPAPVLSFLRPSMDLVSAFYSLAPLSPTSPVSPVSGR 62

DB 109 PPSASAPRRRCPPPPR-----PS---TSALVLLPVSLPSSSSPLIPCSPPR 155

RESULT 8
 T19319
 Hypothetical protein C15H11.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T19319
 R:Bardill, S.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19107
 A:Accession: T19319
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-503 <WIL>
 A:Cross-references: EMBL:281035; PIDN:CAB02737.1; GSPDB:GN00023; CESP:C15H11.5
 A:Experimental source: clone C15H11
 C:Genetics:
 A:Gene: CESP:C15H11.5
 A:Map position: 5
 A:Introns: 53/3; 156/3; 254/3; 292/2; 331/3; 379/3; 437/1

Query Match 18.4%; Score 77; DB 2; Length 503;
 Best Local Similarity 32.7%; Pred. No. 2.6;
 Matches 17; Conservative 8; Mismatches 17; Indels 10; Gaps 1;

OY 11 AVPVPLRMQGPAPVLSFLRPSMDLVSAFYSLAPLSPTSPVSPVSGR 62
 DB 298 SIPLRMQEPG-----SGWYILSAMYSLPVNPVMTGVIRPAEVAR 339

RESULT 9
 T19673
 Hypothetical protein C33B4.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T19673
 R:Coles, L.
 submitted to the EMBL Data Library, February 1995
 A:Reference number: Z19160
 A:Accession: T19673
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1110 <WIL>
 A:Cross-references: EMBL:248367; PIDN:CAA88324.1; GSPDB:GN00020; CESP:C33B4.3
 A:Experimental source: clone C33B4
 C:Genetics:
 A:Gene: CESP:C33B4.3
 A:Map position: 2
 A:Introns: 20/3; 110/3; 144/3; 341/1; 455/1; 546/1; 653/3; 934/2; 983/3; 1035/1

Query Match 18.3%; Score 76.5; DB 2; Length 1110;
 Best Local Similarity 28.6%; Pred. No. 7.3;
 Matches 24; Conservative 12; Mismatches 33; Indels 15; Gaps 4;

OY 3 HSLPAPVPLRMQ-----PGAPVLSFLRPSMDLVSAFYSLAPLSPTSPV 56
 DB 744 HPSLRSASTPQIOQOQSSTPPPPPPPPHCEPT--MNVHETPTSTSSVPPPPPLP 801
 OY 57 PVSVGRGPPD-----AHVAVN 73
 DB 802 PISGAPPPPPPPPGGLMHAAS 825

RESULT 10
 A96826
 T8K14.10 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: A96826
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A96826
 A:Accession: A96826
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-356 <STO>
 A:Cross-references: GB:AE005173; NID:g4835761; PIDN:AA030228.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T8K14.10
 A:Map position: 1

Query Match 18.2%; Score 76; DB 2; Length 356;
 Best Local Similarity 34.7%; Pred. No. 2.2;
 Matches 26; Conservative 7; Mismatches 30; Indels 12; Gaps 3;

OY 9 PAAPVPLRMQGPAPHPVLSFLRPSMDLVSAF-----YSLPLAPLSPTSPISPV---VSQ 60
 DB 216 PCGVLCPPPTSEPPSTPTGTSIPSPS-----SGFLPPIYVPPMAPSPSPVTSATYWCVAK 271
 OY 61 GRGPDPAHVAVNL 75
 DB 272 PSVDPPIIOEAMNFA 286

RESULT 11
 S48273
 probable transcription factor YBR108w - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein YBR0901
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
 C:Accession: S48273; S45976; S44688
 R:Manhaupt, G.; Stucka, R.; Ehme, S.; Vetter, I.; Felmann, H.
 Yeast 10, 1363-1381, 1994
 A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
 A:Reference number: S48255; MUID:95208357
 A:Accession: S48273
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-848 <MAN>
 A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CAAS5611.1; PID:g476064
 R:Felmann, H.; Manhaupt, G.; Schwarzlose, C.; Vetter, I.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45927
 A:Accession: S45976
 A:Molecule type: DNA
 A:Residues: 1-848 <FE2>
 A:Cross-references: EMBL:Z35977; NID:g536378; PID:g536379; MIPS:YBR108w
 C:Genetics:
 A:Map position: 2R

Query Match 18.2%; Score 76; DB 2; Length 848;
 Best Local Similarity 34.3%; Pred. No. 6;
 Matches 23; Conservative 6; Mismatches 28; Indels 10; Gaps 3;

OY 13 PVPRLMQGPAPHPV-----ISFLRPSMDLVSA---FYSPLAPLSPTSPV---ISPVSGR 62
 DB 383 PVPVRMQGPAPHPV-----ISFLRPSMDLVSA---FYSPLAPLSPTSPV---ISPVSGR 62
 OY 63 GRPDPAN 69
 DB 443 LPPPTN 449

RESULT 12
 D49600
 genome-linked protein Vpg - soybean dwarf virus
 C:Species: soybean dwarf virus
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
 C:Accession: D49600
 R:Rathjen, J.P.; Karageorgos, L.E.; Hahill, N.; Waterhouse, P.M.; Symons, R.H.
 Virology 198, 671-679, 1994
 A:Title: Soybean dwarf luteovirus contains the third variant genome type in the luteo
 A:Reference number: A49600; MUID:94120742
 A:Accession: D49600
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-189 <RAU>
 A:Cross-references: GB:L24049; NID:g436017; PIDN:AA17538.1; PID:g436021
 C:Superfamily: potato leaf roll virus genome-linked protein

Query Match 17.9%; Score 75; DB 2; Length 189;
 Best Local Similarity 31.1%; Pred. No. 1.4;
 Matches 19; Conservative 12; Mismatches 22; Indels 8; Gaps 2;

OY 2 THSLPRAAPVPL-----RMQGPAPHPVLSFLRPSMDLVSAF---FYSPLAPLSPTSPV 53
 DB 58 THSCQRTASMVVPPREVSLGRLYONASHSLMEYSRPTWNIKRSVSYSSSRPLPPQV 117
 OY 54 P 54
 DB 118 P 118

RESULT 13
 T48814
 hypothetical protein 15E6.220 [imported] - *Neurospora crassa*
 C:Species: *Neurospora crassa*
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
 C:Accession: T48814
 R:Schlitt, U.; Alqn, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z24541
 A:Accession: T48814
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1952 <SCH>
 A:Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:15E6.220
 A:Experimental source: cosmid contig 15E6; strain 74
 C:Genetics:
 A:Gene: NCSP:15E6.220
 A:Map position: 2
 A:introns: 281/3

Query Match 17.9%; Score 75; DB 2; Length 1952;
 Best Local Similarity 34.8%; Pred. No. 20;
 Matches 24; Conservative 8; Mismatches 31; Indels 6; Gaps 3;

OY 2 THSLPRAAPVPLRMQGPAPHPVLSFLRPSMDLVSAF---FYSPLAPLSPTSPV---ISPV 58
 DB 1781 TPQAVPRPSTALTPTAQGPVSPAVS--GSGVPAASAQSVAPAVSVSTPVPAATVAPA 1838
 OY 59 S-VGRGPD 66
 DB 1839 STVAAAPT 1847

RESULT 14
 T13029
 beta-adaptin homolog F8L21.170 - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Nov-1999
 C:Accession: T13029

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew
submitted to the Protein Sequence Database, July 1999
A:Reference number: Z17587

A:Accession: T13029

A:Molecule type: DNA

A:Residues: 1-894 <BEV>

A:Cross-references: EMBL:AL096882; GSPDB:GN00062; ATSP:F8L21.170

A:Experimental source: cultivar Columbia; BAC clone F8L21

C:Genetics:

A:Gene: ATSP:F8L21.170

A:Map position: 4

A:Introns: 29/3; 95/3; 133/3; 217/2; 262/3; 322/3; 353/3; 424/2; 445/3; 505/3; 536/3; 66
C:Superfamily: beta-adaptin

Query Match 17.8%; Score 74.5; DB 2; Length 894;

Best Local Similarity 29.6%; Pred. No. 9.1;

Matches 24; Conservative 8; Mismatches 30; Indels 19; Gaps 3;

OY 9 PAAVVPRLMQPGPAHPVLSFLRPSMDLVSATFSLPLAPLSPTSVPISE-----PV 58

DB 619 PCNIFQPSGRCQAPAVPA-----PVPDLIGDLMGLDNNAIVPVDPTITQSGPPLPVVPA 673

OY 59 SVGRGPPDPAHVAVNLISRYEG 79

DB 674 SSGQG-----LQISQQLSRKDG 690

RESULT 15

T00273

hypothetical protein KIAA0595 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000

C:Accession: T00273

R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Momura, N.; Ohara, O.

DNA Res. 5, 31-39, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complet

A:Reference number: Z14086; MUID:98290545

A:Accession: T00273

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1520 <NAG>

A:Cross-references: EMBL:AB011167; NID:g3043713; PIDN:BAA25521.1; PID:g3043714

A:Experimental source: brain

C:Genetics:

A:Note: KIAA0595

Query Match 17.5%; Score 73; DB 2; Length 1520;

Best Local Similarity 35.1%; Pred. No. 23;

Matches 27; Conservative 3; Mismatches 25; Indels 22; Gaps 4;

OY 5 LIPRPA-----AVPYPLRMQPGPAHPVLSFLRPSMDLVSATFSLPLAPLSPTSVPISE 56

DB 710 LLARSPPVQSVSPAVPTPPSMSALPFAGGLGMP-----SLPPPLQPPSLPLS 761

OY 57 --PVSVGRGPPDPAHVA 71

DB 762 MGPVL-----PDPTTHYA 774

Search completed: April 11, 2002, 09:30:07
Job time: 104 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:29:13 ; Search time 15.52 Seconds
(without alignments)
186.632 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418
Sequence: 1 GTHSLPRRAAVPVLRMQP.....VGRGPDPAHVAVNLSRYEG 79

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.5	20.0	888	SM6B_HUMAN	Q9H3T3 homo sapien
2	81.5	19.5	1257	PGCN_RAT	P55067 ratius norv
3	80	19.1	428	FXB2_MOUSE	Q64733 mus musculu
4	80	19.1	1840	RPB1_ARATH	P18616 arabidopsis
5	76	18.2	848	YBVB_YEAST	P38266 saccharomyc
6	74.5	17.8	1255	PER2_HUMAN	Q15055 homo sapien
7	73.5	17.6	864	YC18_HUMAN	Q9ULX2 homo sapien
8	73.5	17.6	887	SM6B_RAT	O70141 ratius norv
9	72	17.2	440	NK4R_HUMAN	P30098 homo sapien
10	71.5	17.1	2441	CBP_MOUSE	P45481 mus musculu
11	70.5	16.9	291	YD53_STYNY3	P42350 synechocyst
12	70.5	16.9	788	DPOL_HPBHE	P13846 heron hepat
13	70.5	16.9	1268	PGCN_MOUSE	P35066 mus musculu
14	70	16.7	2318	NTC3_MOUSE	Q61982 mus musculu
15	69.5	16.6	433	TEF5_CHICK	Q90701 gallus gall
16	69	16.6	265	CDX1_HUMAN	P47907 homo sapien
17	68.5	16.4	886	SM6B_MOUSE	O54951 mus musculu
18	68	16.3	703	MU51_HUMAN	Q15231 homo sapien
19	68	16.3	1206	FM14_MOUSE	Q05853 mus musculu
20	68	16.3	1468	FMN1_MOUSE	Q05860 mus musculu
21	68	16.3	1752	RPB1_SCHPO	P36594 schizosacch
22	68	16.3	1859	RPB1_CAEEL	P16356 caenorhabdi
23	67.5	16.1	123	VST1_HEYBU	P29325 hepatitis e
24	67.5	16.1	123	VST1_HEYMY	O04612 hepatitis e
25	67.5	16.1	213	AMEX_BOVIN	P28217 bos taurus
26	67.5	16.1	506	ARSA_MOUSE	P50428 mus musculu
27	67.5	16.1	667	SIX5_MOUSE	P70178 mus musculu
28	67.5	16.1	827	SOM6_MOUSE	P00645 mus musculu
29	67	16.0	360	A2H5_RABIT	P80191 cryocolagus
30	67	16.0	817	VRP1_YEAST	P37370 saccharomyc
31	67	16.0	1433	Y310_HUMAN	O15027 homo sapien
32	66.5	15.9	259	MSP8_EIMAC	P09125 elmeria ace
33	66.5	15.9	283	EXTN_SORBI	P24152 sorghum bic

34	66.5	15.9	435	1	TEF5_HUMAN	Q99594 homo sapien
35	66.5	15.9	1229	1	N121_HUMAN	O9Y2N3 homo sapien
36	66	15.8	534	1	APG_ARATH	P40602 arabidopsis
37	66	15.8	538	1	TF65_CHICK	P98152 gallus gall
38	66	15.8	868	1	NRG2_RAT	O35569 ratius norv
39	66	15.8	1447	1	DOC_HUMAN	P43146 homo sapien
40	66	15.8	1447	1	DOC_MOUSE	P70211 mus musculu
41	66	15.8	3149	1	TEGU_EBV	P03186 epstein-bar
42	65.5	15.7	357	1	GDF1_MOUSE	P20863 mus musculu
43	65.5	15.7	1257	1	PER2_MOUSE	O54943 mus musculu
44	65.5	15.7	2805	1	MAPA_HUMAN	P78559 homo sapien
45	65	15.6	582	1	MNT_HUMAN	Q99583 homo sapien

ALIGNMENTS

RESULT	ID	SM6B_HUMAN	STANDARD	PRT	888 AA.
AC	Q9H3T3	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	SEMAPHORIN 6B PRECURSOR (SEMAPHORIN Z) (SEMA Z).				
GN	SEMA6B OR SEMAZ.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RA	Kimura T., Ishida H.;				
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.				
CC	- FUNCTION: MAY PLAY A ROLE IN BOTH PERIPHERAL AND CENTRAL NERVOUS				
CC	SYSTEM DEVELOPMENT (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				
CC	- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 SEMA DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AB022433; BAB20669.1; .				
DR	InterPro; IPR003659; PSI.				
DR	InterPro; IPR001627; Sema.				
DR	InterPro; IPR000737; Squash.				
DR	Pfam; PF01403; Sema; 1.				
DR	SMART; SM00423; PSI; 1.				
DR	SMART; SM00286; PTI; 1.				
KW	Signal. Transmembrane; Multi-pass type I; Neurogenesis; Glycoprotein;				
KW	Developmental protein.				
FT	SIGNAL	1	25		POTENTIAL.
FT	CHAIN	26	888		SEMAPHORIN 6B.
FT	DOMAIN	26	603		EXTRACELLULAR (POTENTIAL).
FT	TRANSHEM	604	624		POTENTIAL.
FT	DOMAIN	625	888		CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	238	547		SEMA.
FT	DOMAIN	661	674		POLY-GLY.
FT	DOMAIN	750	753		POLY-LEU.
FT	CARBOHYD	74	74		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	155	155		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	167	167		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	291	291		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	386	386		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	441	441		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	462	462		N-LINKED (GLCNAC. . .) (POTENTIAL).

DB 670 SVPEQAVRPVSG-AEDPE 688

```

RESULT 3
ID FBX2_MOUSE STANDARD: PRT: 428 AA.
AC 064733:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FORKHEAD BOX PROTEIN B2 (TRANSCRIPTION FACTOR FKX-4).
GN FOXB2 OR FKX4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97014266; PubMed=8611101;
RA Kaestner K.H., Schuetz G., Monaghan A.P.;
RT "Expression of the winged helix genes Fkh-4 and Fkh-5 defines domains
RL in the central nervous system.";
RN Mech. Dev. 55:221-230(1996).
RN [2]
RP SEQUENCE OF 4-114 FROM N.A.
RX STRAIN=129;
RA MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemlsch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally
RT regulated.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
RN [1]
RP SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X92591; CA63335.1; -
DR EMBL: X71942; CA50744.1; -
DR TRANSFAC: T02442; -
DR MGD: MGI:1347468; Foxd2.
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head.1
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FKH.1
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS00659; FORK_HEAD_3; 1.
DR PROSITE: PS00659; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear Protein; Transcription regulation.
FT DNAS_BIND 12 103
FT DNAS_BIND 139 153
FT DNAS_BIND 153 153
FT DNAS_BIND 156 162
FT DNAS_BIND 163 172
FT DNAS_BIND 172 172
FT DNAS_BIND 217 231
FT DNAS_BIND 249 258
FT DNAS_BIND 321 330
FT DNAS_BIND 330 330
FT DNAS_BIND 396 399
FT DNAS_BIND 428 AA; 45170 MM; DB8ABEF1E94AB10 CRC64;
SQ SEQUENCE

```

Query Match 19.1%; Score 80; DB 1; Length 428;
 Best Local Similarity 34.8%; Pred. No. 0.66;
 Matches 23; Conservative 10; Mismatches 25; Indels 8; Gaps 2;
 3 HSLPRPAVAVPPLRMGCGAPVPLSLRPSMDLVSAFSLP-----LAPLSPSVIS 56

DB 350 HSAOGLPAVAVPIK--PTPALPVYTTLPALSVPTASQGLPASTVCAAAASPTAPLLE 407

OY 57 PVSAGR 62
 DB 408 PTAAGR 413

```

RESULT 4
ID RPB1_ARATH STANDARD: PRT: 1840 AA.
AC P18616; P31635; O952S8;
DT 01-NOV-1990 (Rel. 16, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (RC 2.7.7.6).
GN RPB205 OR RPII OR RPB1 OR AT4G35800 OR F4B14.70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, COLUMBIA;
RX MEDLINE=91080867; PubMed=2259344;
RA Nawrath C., Scheil J., Koncz C.;
RT "Homologous domains of the largest subunit of eucaryotic RNA
RT polymerase II are conserved in plants.";
RN Mol. Gen. Genet. 223:65-75(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, COLUMBIA;
RX MEDLINE=91355869; PubMed=2103447;
RA Dietrich M.A., Prenger J.P., Guilfoyle T.J.;
RT "Analysis of the genes encoding the largest subunit of RNA polymerase
RT II in Arabidopsis and soybean.";
RN Plant Mol. Biol. 15:207-223(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohensei J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandebussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weltzenegger T., Bothé G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkes W.,
RA Moolman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
RA Berneiser S., Hempel S., Felipausch M., Lambert S., Van den Daele H.,
RA de Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,
RA Portet A., Rajandream M.-A., Lyne M., Benes V., Rehmann T.-H.,
RA Borkova D., Bloedrean M.-A., Scharfe M., Grimm M., Loehert T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Grandtath K., Dauner D., Hezli A.,
RA Neumann S., Argitlou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clapaud G., Muendlein A., Felber R.,
RA Schenabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bietke C.,
RA Fritsman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Beyer M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,

```

RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Mink P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Krimer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Speth J., Ryan E., Andrews S., Gelsel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Dron C., Cotton M., Joshi C.,
RA Antonolu B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shan R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martensen R., McCombie W.R.,
RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana";
RT
RL Nature 402:769-777(1999).
CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -I- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -I- SUBCELLULAR LOCATION: NUCLEAR
CC -I- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -I- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52954; CA37130.1; -;
DR EMBL: X52954; CA37130.1; -;
DR EMBL: AL031986; CA21466.2; -;
DR EMBL: AL161588; CAB81489.1; -;
DR PIR: S12071; JDM01.
DR PIR: S11960; JDM02.
DR InterPro: IPR002965; P-rich-extension.
DR InterPro: IPR000684; RNA_pol.II_repeat.
DR InterPro: IPR000722; RNA_pol.A.
DR InterPro: IPR002879; RNA_pol.A2.
DR Pfam: PF00623; RNA_pol_A; 1.
DR Pfam: PF01854; RNA_pol_A2; 1.
DR PROSITE: PS00115; RNA_POL.II_REPEAT; 23.
KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT ZN_FING 66 79
FT DNA_BIND 327 398
FT BT SIMILARITY
FT DOMAIN 786 796
FT DOMAIN 1530 1819
FT CONFLICT 117 124
FT CONFLICT 193 197
FT CONFLICT 298 298
FT CONFLICT 303 303
FT CONFLICT 401 417
FT CONFLICT 428 428
FT CONFLICT 446 446
FT CONFLICT 739 739
FT CONFLICT 1062 1062
FT CONFLICT 1089 1089
FT CONFLICT 1720 1720
SO SEQUENCE 1840 AA; 204688 MW; 8453621AD945C1B6 CRC64;
Query Match 19.1%; Score 80; DB 1; Length 1840;

Best Local Similarity 36.4%; Pred. No. 3.1;
Matches 28; Conservative 3; Mismatches 36; Indels 10; Gaps 3;
OY 7 PRPAVPPVLMQP-----GPAHPLSLRPSWDLVSAFTSLPLAPLSPTSPVSPVSGR 62
Db 1736 PSIAVPSNARLSASPSPSPVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSSGA 1791
OY 63 GPD--PDHVAVNLRSY 77
Db 1792 SPDVSPSAGYSPPLPGY 1808
RESULT 5
YB8_YEAST STANDARD; PRT; 848 AA.
ID YB8_YEAST
AC P38266;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHEICAL 92.8 KDA PROTEIN IN PHO88-CMD1 INTERGENIC REGION.
GN YBR108W OR YBR0901.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95208357; PubMed=7900426;
RA Mannhaupt G., Stucke R., Ehle S., Vetter I., Feldmann H.;
RT "Analysis of a 70 kb region on the right arm of yeast chromosome II";
RL Yeast 10:1363-1381(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X78993; CA55611.1; -;
DR EMBL: Z35977; CA485063.1; -;
DR PIR: S44688; S44688.
DR SGD: S0000312; YBR108W.
KW Hypothetical protein.
KW SEQUENCE 848 AA; 92762 MW; F33D371369PBAF97 CRC64;
Query Match 18.2%; Score 76; DB 1; Length 848;
Best Local Similarity 34.3%; Pred. No. 3.4;
Matches 23; Conservative 6; Mismatches 28; Indels 10; Gaps 3;
OY 13 PVPVLMQGPAPHPV-----LSFLRPSWDLVSA---FYSPLAPLSPTSPV---ISPVYGR 62
Db 383 PVPVLMQGPAPHPV-----LSFLRPSWDLVSA---FYSPLAPLSPTSPV---ISPVYGR 62
OY 63 GPDPAH 69
Db 443 LPPEPT 449
RESULT 6
PER2_HUMAN STANDARD; PRT; 1255 AA.
ID PER2_HUMAN
AC O15053;
DT 15-JUL-1999 (Rel. 38, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PERIOD CIRCADIAN PROTEIN 2.
GN PER2 OR KIA0347.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 RN [2]
 RP REVISIONS TO C-TERMINUS.
 RC TISSUE=Brain;
 RA Nagase T., Ishikawa K.-I., Seki N., Nakajima D., Ohira M.,
 Miyajima N., Kotani H., Nomura N., Ohara O.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP TISSUE EXPRESSION.
 RX MEDLINE=98087121; PubMed=9427249;
 RA Shearman L.P., Zylka M.J., Weaver D.R., Kolakowski L.F. Jr.,
 Reppert S.M.;
 RT "Two period homologs: circadian expression and photic regulation in
 the suprachiasmatic nuclei.";
 RL Neuron 19:1261-1269(1997).
 CC -1- FUNCTION: CIRCADIAN REGULATOR THAT MAY ACT AS A TRANSCRIPTION
 FACTOR. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN TRANSCRIPTIONAL
 LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING INDIRECT
 TRANSCRIPTIONAL INHIBITION. EXPRESSION OSCILLATES WITH A 24 HR
 RHYTHM IN THE SUPRACHIASMATIC NUCLEUS (SCN) AND THE WHOLE EYES.
 OSCILLATIONS ARE MAINTAINED UNDER CONSTANT DARKNESS AND ARE
 RESPONSIVE TO CHANGES OF THE LIGHT/DARK CYCLES. THERE IS A 4 HOUR
 TIME DELAY BETWEEN PER1 AND PER2 OSCILLATIONS. THE EXPRESSION
 RHYTHMS APPEAR TO ORIGINATE FROM RETINA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND IN HEART, BRAIN,
 PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS.
 CC -1- INDUCTION: BY LIGHT (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 TRANSCRIPTION FACTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch)
 CC -----
 DR EMBL; AB002345; BAA20804.2; ALT_INIT.
 DR MIM; 603426; -
 DR InterPro; IPR001610; PAC.
 DR InterPro; IPR000014; PAS.
 DR SMART; SM00086; PAC; 1.
 DR SMART; SM00091; PAS; 2.
 KM Transcription regulation; Nuclear protein; Repeat; Biological rhythms.
 FT DOMAIN 109 146 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT REPEAT 182 248 PAS-1.
 FT REPEAT 322 386 PAS-2.
 FT DOMAIN 398 438 PAC MOTIF.
 FT DOMAIN 510 513 POLY-ARG.
 FT DOMAIN 789 806 NUCLEAR LOCALIZATION SIGNAL.
 FT DOMAIN 842 979 PRO-RICH.
 SO SEQUENCE 1255 AA; 136579 MW; 2AEF2C6BD4B6CBB0 CRC64;

Query Match 17.8%; Score 74.5; DB 1; Length 1255;
 Best Local Similarity 25.6%; Pred. No. 7.2;
 Matches 30; Conservative 10; Mismatches 26; Indels 51; Gaps 5;
 Oy 1 GTHSLPRP-----AAVPLRLMQ-----PGPAHPVLSFLRPSMDVLSAFYSILPL 45

Db 863 GTVAAPAPAPPHASFVPAVPVLDLQHFAVOPPPFPAPLAPVMAFWLPSFSFGPNLPQ 922
 Oy 46 A-PLSP-----TSVP-----ISVSVGRGEPD 66
 Db 923 AFPPSQPQFPSPHPTLTSEMASASQPEFSPRTSIPQPCACPATRTATPPSAMGRASP 979
 RESULT 7
 YC18_HUMAN STANDARD; PRT; 864 AA.
 AC Q90LK2;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN KIAA1218 (FRAGMENT).
 GN KIAA1218.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
 Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XV.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 6:337-345(1999).
 CC -1- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch)
 CC -----
 DR EMBL; AB033044; BAA6532.1; -
 KW Hypothetical protein.
 FT NON_TER 1
 FT DOMAIN 223 226 POLY-THR.
 FT DOMAIN 652 664 POLY-SER.
 FT DOMAIN 786 790 POLY-SER.
 SO SEQUENCE 864 AA; 92138 MW; D55CB8130E48DA23 CRC64;

Query Match 17.6%; Score 73.5; DB 1; Length 864;
 Best Local Similarity 35.5%; Pred. No. 6.1;
 Matches 33; Conservative 4; Mismatches 21; Indels 35; Gaps 7;
 Oy 4 SLPRPA--VVPPLR-MOPGAHPVLSFLRPSMDVLSAFYSILPLSP-----T 51
 Db 517 SPLSPRAHHTTPVPASVLP-----FSNP-----SAVY-LPSADISSRLTSSITMT 562
 Oy 52 SVPISPVSVGRGPPDA-----HYAVNLVS 75
 Db 563 SAMLSTNAFVTSPPSALMSHTTAFPHVAATLS 595
 RESULT 8
 SM6B_RAT STANDARD; PRT; 887 AA.
 AC 070141;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SEMAPHORIN 6B PRECURSOR (SEMAPHORIN Z) (SEMA Z).
 GN SEMA6B.
 OS Rattus norvegicus (Rat).

Query Match 17.2% Score 72; DB 1; Length 440;
 Best Local Similarity 45.2% Pred. No. 4.2; Mismatches 10; Indels 8; Gaps 3;
 Matches 19; Conservative 5; Mismatches 10; Indels 8; Gaps 3;

2 THSLPRAVAVPLRMOPGPAHVLSEFLRPSMDLVSAFYSL 43
 32 TASPSPAPSWTSPSP---RPGPAHPP---FLQPPMAV---ALMSL 65

RESULT 10
 CBP_MOUSE
 ID CBP_MOUSE STANDARD; PRT; 2441 AA.
 AC P45481;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-FEB-1996 (rel. 33, Last sequence update)
 DT 15-JUL-1998 (rel. 36, Last annotation update)
 DE CREB-BINDING PROTEIN.
 GN CREBBP OR CBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE=94019866; PubMed=8413673;
 RA Christia J.C., Kwok R.P.S., Lamb N., Hagihara M., Montminy M.R.,
 RA Goodman R.H.;
 RT "Phosphorylated CREB binds specifically to the nuclear protein CBP.";
 RL Nature 365:855-859(1993).
 CC - FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
 PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
 THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
 CAMP-RESPONSIVE GENES.
 CC - SUBCELLULAR LOCATION: NUCLEAR.
 CC - SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S66385; AAB28651.1; -;
 DR TRANSFAC; T01318; -;
 DR MGD; MGI:1098280; Crebbp.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR003101; KIX.
 DR InterPro; IPR000197; TAF_finger.
 DR InterPro; IPR000433; ZnF_Z2.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF02172; KIX; 1.
 DR Pfam; PF02135; zf-TAF; 2.
 DR Pfam; PF00569; Z2; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00291; ZNF_Z2; 1.
 DR SMART; SM00291; BROMO; 1.
 DR PROSITE; PS00633; BROMODOMAIN.1; 1.
 DR PROSITE; PS50014; BROMODOMAIN.2; 1.
 KW Transcription regulation; Nuclear protein; Activator; Bromodomain.
 FT DOMAIN 1104 1176 BROMODOMAIN.
 FT DOMAIN 1062 1065 POLY-GLU.
 FT DOMAIN 1556 1563 POLY-GLU.
 FT DOMAIN 1944 1949 POLY-PRO.
 FT DOMAIN 1968 1971 POLY-GLN.
 FT DOMAIN 2082 2086 POLY-GLN.
 FT DOMAIN 2200 2216 POLY-GLN.
 FT DOMAIN 2296 2299 POLY-GLN.
 FT SEQUENCE 2441 AA; 265474 MW; 0ABB028C3112FA19 CRC64;

Query Match 17.1% Score 71.5; DB 1; Length 2441;
 Best Local Similarity 36.5% Pred. No. 28; Mismatches 33; Indels 3; Gaps 2;
 Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

4 SLPLPRAVAVPLRMOPGPAHVLSEFLRPSMDLVSAFYSLPLAPSPVSPISPVSGRG 63
 843 SQLPCRPVQSPFLHPPPPASTAAGN--PSLQHPTAPGHTPPQAPAPIQ--PSTPVSSGOT 899

QY 64 PDP 66
 Db 900 PTP 902

RESULT 11
 YD53_SYNY3
 ID YD53_SYNY3 STANDARD; PRT; 291 AA.
 AC P42350;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE HYPOTHETICAL 31.3 KDA PROTEIN SLR1353.
 GN SLR1353.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxId=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9322488; PubMed=8467083;
 RA Malakhov M.P., Wada H., Los D.A., Sakamoto T., Murata N.;
 RT "Structure of a cyanobacterial gene encoding the 50S ribosomal
 protein L9.";
 RL Plant Mol. Biol. 21:913-918(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugiyura M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
 RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10716; BAA38818.1; -;
 DR EMBL; D90912; BAA18174.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 291 AA; 31283 MW; 5AB7E3DD03C36390 CRC64;

Query Match 16.9% Score 70.5; DB 1; Length 291;
 Best Local Similarity 28.8% Pred. No. 3.8; Mismatches 30; Indels 11; Gaps 5;
 Matches 23; Conservative 16; Mismatches 30; Indels 11; Gaps 5;

4 SLPLPRAVAVPLRMOPGPAHVLSEFLRPSMD--LVSAFYSLP--LAPLSPT-----SVP 54
 145 AIAPEPTLTPAPISPPSP--DVLSTLEPTTPPPAVNVSFNQPPESAPIDSELQIDFAFP 203

QY 55 ISPVSYGRGPD--PDAHVAVN 73
 Db 204 ELPLAVEARPDSPDPMAVS 223


```

FT DISULFID 969 984 BY SIMILARITY.
FT DISULFID 986 995 BY SIMILARITY.
FT DISULFID 1040 1051 BY SIMILARITY.
FT DISULFID 1068 1160 BY SIMILARITY.
FT DISULFID 1136 1152 BY SIMILARITY.
FT DISULFID 1167 1210 BY SIMILARITY.
FT DISULFID 1196 1223 BY SIMILARITY.
FT CAROAMD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROAMD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROAMD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROAMD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROAMD 1175 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1368 AA; 137200 MW; 3014E8E20A2FAEC CRC64;

Query Match 16.9%; Score 70.5; DB 1; Length 1268;
Best Local Similarity 26.1%; Pred. No. 18;
Matches 30; Conservative 11; Mismatches 19; Indels 55; Gaps 7;

OY 6 LPR-PAAVVPLRMQPGPAH-----VLSFLR-----PSMDLYSA 39
DB 604 LPRLSSEPPAP---SPGSEALSANSLQSSADGSPDFIVAMLRPKMLPRSTLVN 660
OY 40 FYSLPLAPLS--PTSV--ISPVSVG-----RGPPDDA 68
DB 661 MTPVPLSPASPLSMVPEQAVRVSIGADELTPQTITAPVEASHRSPDADS 715

RESULT 14
NTC3_MOUSE STANDARD; PRT; 2318 AA.
AC 061982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE "NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN NOTCH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR X SWISS WEBSTER;
RX MEDLINE=95001556; PubMed=7918097;
RA Lardelli M., Dalestrand J., Lendahl U.;
RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT growth factor-repeats and is expressed in proliferating
RT neuroepithelium".
RL Mech. Dev. 46:123-136(1994).
CC -1- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC -1- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC -1- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
-----
CC EMBL; X74760; CAA52776.1;
CC HSSP; P00740; IIXA.
DR MGD; MGI:99460; Notch3.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF000023; ank; 6.
DR Pfam; PF00008; EGF; 34.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBLD.
DR SMART; SM00178; ANK; 5.
DR SMART; SM00001; EGF_Ca; 19.
DR SMART; SM00004; EGF_Like; 15.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 18.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS01187; EGF_Ca; 17.
KW Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
KW ANK repeat; Glycoprotein.
FT 1 1643
FT 2 1644
FT 3 1664
FT 4 1665
FT 5 2318
FT 6 2318
FT 7 2242
FT 8 2261
FT 9 78
FT 10 39
FT 11 79
FT 12 119
FT 13 157
FT 14 159
FT 15 196
FT 16 198
FT 17 235
FT 18 237
FT 19 273
FT 20 313
FT 21 315
FT 22 351
FT 23 352
FT 24 390
FT 25 392
FT 26 432
FT 27 432
FT 28 470
FT 29 508
FT 30 544
FT 31 546
FT 32 581
FT 33 583
FT 34 621
FT 35 656
FT 36 694
FT 37 696
FT 38 731
FT 39 735
FT 40 771
FT 41 772
FT 42 809
FT 43 811
FT 44 848
FT 45 850
FT 46 888
FT 47 923
FT 48 925
FT 49 961
FT 50 963
FT 51 1001
FT 52 1035
FT 53 1037
FT 54 1085
FT 55 1121
FT 56 1123
FT 57 1159
FT 58 1204
FT 59 1206
FT 60 1245
FT 61 1247
FT 62 1288
FT 63 1290
FT 64 1326
FT 65 1336
FT 66 1374
FT 67 1388
FT 68 1428
FT 69 1429
FT 70 1467
FT 71 1468
FT 72 1503
FT 73 1503
FT 74 1839
FT 75 1868
FT 76 1872
FT 77 1902
FT 78 1906
FT 79 1935
FT 80 1938
FT 81 1939
FT 82 1968
FT 83 1972
FT 84 2001
FT 85 43
FT 86 49
FT 87 55
FT 88 66
FT 89 77
FT 90 83
FT 91 94
FT 92 107
FT 93 107
FT 94 118
FT 95 118
FT 96 124
FT 97 124
FT 98 135
FT 99 135
FT 100 135
FT 101 135
FT 102 135
FT 103 135
FT 104 135
FT 105 135
FT 106 135
FT 107 135
FT 108 135
FT 109 135
FT 110 135
FT 111 135
FT 112 135
FT 113 135
FT 114 135
FT 115 135
FT 116 135
FT 117 135
FT 118 135
FT 119 135
FT 120 135
FT 121 135
FT 122 135
FT 123 135
FT 124 135
FT 125 135
FT 126 135
FT 127 135
FT 128 135
FT 129 135
FT 130 135
FT 131 135
FT 132 135
FT 133 135
FT 134 135
FT 135 135
FT 136 135
FT 137 135
FT 138 135
FT 139 135
FT 140 135
FT 141 135
FT 142 135
FT 143 135
FT 144 135
FT 145 135
FT 146 135
FT 147 135
FT 148 135
FT 149 135
FT 150 135
FT 151 135
FT 152 135
FT 153 135
FT 154 135
FT 155 135
FT 156 135
FT 157 135
FT 158 135
FT 159 135
FT 160 135
FT 161 135
FT 162 135
FT 163 135
FT 164 135
FT 165 135
FT 166 135
FT 167 135
FT 168 135
FT 169 135
FT 170 135
FT 171 135
FT 172 135
FT 173 135
FT 174 135
FT 175 135
FT 176 135
FT 177 135
FT 178 135
FT 179 135
FT 180 135
FT 181 135
FT 182 135
FT 183 135
FT 184 135
FT 185 135
FT 186 135
FT 187 135
FT 188 135
FT 189 135
FT 190 135
FT 191 135
FT 192 135
FT 193 135
FT 194 135
FT 195 135
FT 196 135
FT 197 135
FT 198 135
FT 199 135
FT 200 135
FT 201 135
FT 202 135
FT 203 135
FT 204 135
FT 205 135
FT 206 135
FT 207 135
FT 208 135
FT 209 135
FT 210 135
FT 211 135
FT 212 135
FT 213 135
FT 214 135
FT 215 135
FT 216 135
FT 217 135
FT 218 135
FT 219 135
FT 220 135
FT 221 135
FT 222 135
FT 223 135
FT 224 135
FT 225 135
FT 226 135
FT 227 135
FT 228 135
FT 229 135
FT 230 135
FT 231 135
FT 232 135
FT 233 135
FT 234 135
FT 235 135
FT 236 135
FT 237 135
FT 238 135
FT 239 135
FT 240 135
FT 241 135
FT 242 135
FT 243 135
FT 244 135
FT 245 135
FT 246 135
FT 247 135
FT 248 135
FT 249 135
FT 250 135
FT 251 135
FT 252 135
FT 253 135
FT 254 135
FT 255 135
FT 256 135
FT 257 135
FT 258 135
FT 259 135
FT 260 135
FT 261 135
FT 262 135
FT 263 135
FT 264 135
FT 265 135
FT 266 135
FT 267 135
FT 268 135
FT 269 135
FT 270 135
FT 271 135
FT 272 135
FT 273 135
FT 274 135
FT 275 135
FT 276 135
FT 277 135
FT 278 135
FT 279 135
FT 280 135
FT 281 135
FT 282 135
FT 283 135
FT 284 135
FT 285 135
FT 286 135
FT 287 135
FT 288 135
FT 289 135
FT 290 135
FT 291 135
FT 292 135
FT 293 135
FT 294 135
FT 295 135
FT 296 135
FT 297 135
FT 298 135
FT 299 135
FT 300 135
FT 301 135
FT 302 135
FT 303 135
FT 304 135
FT 305 135
FT 306 135
FT 307 135
FT 308 135
FT 309 135
FT 310 135
FT 311 135
FT 312 135
FT 313 135
FT 314 135
FT 315 135
FT 316 135
FT 317 135
FT 318 135
FT 319 135
FT 320 135
FT 321 135
FT 322 135
FT 323 135
FT 324 135
FT 325 135
FT 326 135
FT 327 135
FT 328 135
FT 329 135
FT 330 135
FT 331 135
FT 332 135
FT 333 135
FT 334 135
FT 335 135
FT 336 135
FT 337 135
FT 338 135
FT 339 135
FT 340 135
FT 341 135
FT 342 135
FT 343 135
FT 344 135
FT 345 135
FT 346 135
FT 347 135
FT 348 135
FT 349 135
FT 350 135
FT 351 135
FT 352 135
FT 353 135
FT 354 135
FT 355 135
FT 356 135
FT 357 135
FT 358 135
FT 359 135
FT 360 135
FT 361 135
FT 362 135
FT 363 135
FT 364 135
FT 365 135
FT 366 135
FT 367 135
FT 368 135
FT 369 135
FT 370 135
FT 371 135
FT 372 135
FT 373 135
FT 374 135
FT 375 135
FT 376 135
FT 377 135
FT 378 135
FT 379 135
FT 380 135
FT 381 135
FT 382 135
FT 383 135
FT 384 135
FT 385 135
FT 386 135
FT 387 135
FT 388 135
FT 389 135
FT 390 135
FT 391 135
FT 392 135
FT 393 135
FT 394 135
FT 395 135
FT 396 135
FT 397 135
FT 398 135
FT 399 135
FT 400 135
FT 401 135
FT 402 135
FT 403 135
FT 404 135
FT 405 135
FT 406 135
FT 407 135
FT 408 135
FT 409 135
FT 410 135
FT 411 135
FT 412 135
FT 413 135
FT 414 135
FT 415 135
FT 416 135
FT 417 135
FT 418 135
FT 419 135
FT 420 135
FT 421 135
FT 422 135
FT 423 135
FT 424 135
FT 425 135
FT 426 135
FT 427 135
FT 428 135
FT 429 135
FT 430 135
FT 431 135
FT 432 135
FT 433 135
FT 434 135
FT 435 135
FT 436 135
FT 437 135
FT 438 135
FT 439 135
FT 440 135
FT 441 135
FT 442 135
FT 443 135
FT 444 135
FT 445 135
FT 446 135
FT 447 135
FT 448 135
FT 449 135
FT 450 135
FT 451 135
FT 452 135
FT 453 135
FT 454 135
FT 455 135
FT 456 135
FT 457 135
FT 458 135
FT 459 135
FT 460 135
FT 461 135
FT 462 135
FT 463 135
FT 464 135
FT 465 135
FT 466 135
FT 467 135
FT 468 135
FT 469 135
FT 470 135
FT 471 135
FT 472 135
FT 473 135
FT 474 135
FT 475 135
FT 476 135
FT 477 135
FT 478 135
FT 479 135
FT 480 135
FT 481 135
FT 482 135
FT 483 135
FT 484 135
FT 485 135
FT 486 135
FT 487 135
FT 488 135
FT 489 135
FT 490 135
FT 491 135
FT 492 135
FT 493 135
FT 494 135
FT 495 135
FT 496 135
FT 497 135
FT 498 135
FT 499 135
FT 500 135
FT 501 135
FT 502 135
FT 503 135
FT 504 135
FT 505 135
FT 506 135
FT 507 135
FT 508 135
FT 509 135
FT 510 135
FT 511 135
FT 512 135
FT 513 135
FT 514 135
FT 515 135
FT 516 135
FT 517 135
FT 518 135
FT 519 135
FT 520 135
FT 521 135
FT 522 135
FT 523 135
FT 524 135
FT 525 135
FT 526 135
FT 527 135
FT 528 135
FT 529 135
FT 530 135
FT 531 135
FT 532 135
FT 533 135
FT 534 135
FT 535 135
FT 536 135
FT 537 135
FT 538 135
FT 539 135
FT 540 135
FT 541 135
FT 542 135
FT 543 135
FT 544 135
FT 545 135
FT 546 135
FT 547 135
FT 548 135
FT 549 135
FT 550 135
FT 551 135
FT 552 135
FT 553 135
FT 554 135
FT 555 135
FT 556 135
FT 557 135
FT 558 135
FT 559 135
FT 560 135
FT 561 135
FT 562 135
FT 563 135
FT 564 135
FT 565 135
FT 566 135
FT 567 135
FT 568 135
FT 569 135
FT 570 135
FT 571 135
FT 572 135
FT 573 135
FT 574 135
FT 575 135
FT 576 135
FT 577 135
FT 578 135
FT 579 135
FT 580 135
FT 581 135
FT 582 135
FT 583 135
FT 584 135
FT 585 135
FT 586 135
FT 587 135
FT 588 135
FT 589 135
FT 590 135
FT 591 135
FT 592 135
FT 593 135
FT 594 135
FT 595 135
FT 596 135
FT 597 135
FT 598 135
FT 599 135
FT 600 135
FT 601 135
FT 602 135
FT 603 135
FT 604 135
FT 605 135
FT 606 135
FT 607 135
FT 608 135
FT 609 135
FT 610 135
FT 611 135
FT 612 135
FT 613 135
FT 614 135
FT 615 135
FT 616 135
FT 617 135
FT 618 135
FT 619 135
FT 620 135
FT 621 135
FT 622 135
FT 623 135
FT 624 135
FT 625 135
FT 626 135
FT 627 135
FT 628 135
FT 629 135
FT 630 135
FT 631 135
FT 632 135
FT 633 135
FT 634 135
FT 635 135
FT 636 135
FT 637 135
FT 638 135
FT 639 135
FT 640 135
FT 641 135
FT 642 135
FT 643 135
FT 644 135
FT 645 135
FT 646 135
FT 647 135
FT 648 135
FT 649 135
FT 650 135
FT 651 135
FT 652 135
FT 653 135
FT 654 135
FT 655 135
FT 656 135
FT 657 135
FT 658 135
FT 659 135
FT 660 135
FT 661 135
FT 662 135
FT 663 135
FT 664 135
FT 665 135
FT 666 135
FT 667 135
FT 668 135
FT 669 135
FT 670 135
FT 671 135
FT 672 135
FT 673 135
FT 674 135
FT 675 135
FT 676 135
FT 677 135
FT 678 135
FT 679 135
FT 680 135
FT 681 135
FT 682 135
FT 683 135
FT 684 135
FT 685 135
FT 686 135
FT 687 135
FT 688 135
FT 689 135
FT 690 135
FT 691 135
FT 692 135
FT 693 135
FT 694 135
FT 695 135
FT 696 135
FT 697 135
FT 698 135
FT 699 135
FT 700 135
FT 701 135
FT 702 135
FT 703 135
FT 704 135
FT 705 135
FT 706 135
FT 707 135
FT 708 135
FT 709 135
FT 710 135
FT 711 135
FT 712 135
FT 713 135
FT 714 135
FT 715 135
FT 716 135
FT 717 135
FT 718 135
FT 719 135
FT 720 135
FT 721 135
FT 722 135
FT 723 135
FT 724 135
FT 725 135
FT 726 135
FT 727 135
FT 728 135
FT 729 135
FT 730 135
FT 731 135
FT 732 135
FT 733 135
FT 734 135
FT 735 135
FT 736 135
FT 737 135
FT 738 135
FT 739 135
FT 740 135
FT 741 135
FT 742 135
FT 743 135
FT 744 135
FT 745 135
FT 746 135
FT 747 135
FT 748 135
FT 749 135
FT 750 135
FT 751 135
FT 752 135
FT 753 135
FT 754 135
FT 755 135
FT 756 135
FT 757 135
FT 758 135
FT 759 135
FT 760 135
FT 761 135
FT 762 135
FT 763 135
FT 764 135
FT 765 135
FT 766 135
FT 767 135
FT 768 135
FT 769 135
FT 770 135
FT 771 135
FT 772 135
FT 773 135
FT 774 135
FT 775 135
FT 776 135
FT 777 135
FT 778 135
FT 779 135
FT 780 135
FT 781 135
FT 782 135
FT 783 135
FT 784 135
FT 785 135
FT 786 135
FT 787 135
FT 788 135
FT 789 135
FT 790 135
FT 791 135
FT 792 135
FT 793 135
FT 794 135
FT 795 135
FT 796 135
FT 797 135
FT 798 135
FT 799 135
FT 800 135
FT 801 135
FT 802 135
FT 803 135
FT 804 135
FT 805 135
FT 806 135
FT 807 135
FT 808 135
FT 809 135
FT 810 135
FT 811 135
FT 812 135
FT 813 135
FT 814 135
FT 815 135
FT 816 135
FT 817 135
FT 818 135
FT 819 135
FT 820 135
FT 821 135
FT 822 135
FT 823 135
FT 824 135
FT 825 135
FT 826 135
FT 827 135
FT 828 135
FT 829 135
FT 830 135
FT 831 135
FT 832 135
FT 833 135
FT 834 135
FT 835 135
FT 836 135
FT 837 135
FT 838 135
FT 839 135
FT 840 135
FT 841 135
FT 842 135
FT 843 135
FT 844 135
FT 845 135
FT 846 135
FT 847 135
FT 848 135
FT 849 135
FT 850 135
FT 851 135
FT 852 135
FT 853 135
FT 854 135
FT 855 135
FT 856 135
FT 857 135
FT 858 135
FT 859 135
FT 860 135
FT 861 135
FT 862 135
FT 863 135
FT 864 135
FT 865 135
FT 866 135
FT 867 135
FT 868 135
FT 869 135
FT 870 135
FT 871 135
FT 872 135
FT 873 135
FT 874 135
FT 875 135
FT 876 135
FT 877 135
FT 878 135
FT 879 135
FT 880 135
FT 881 135
FT 882 135
FT 883 135
FT 884 135
FT 885 135
FT 886 135
FT 887 135
FT 888 135
FT 889 135
FT 890 135
FT 891 135
FT 892 135
FT 893 135
FT 894 135
FT 895 135
FT 896 135
FT 897 135
FT 898 135
FT 899 135
FT 900 135
FT 901 135
FT 902 135
FT 903 135
FT 904 135
FT 905 135
FT 906 135
FT 907 135
FT 908 135
FT 909 135
FT 910 135
FT 911 135
FT 912 135
FT 913 135
FT 914 135
FT 915 135
FT 916 135
FT 917 135
FT 918 135
FT 919 135
FT 920 135
FT 921 135
FT 922 135
FT 923 135
FT 924 135
FT 925 135
FT 926 135
FT 927 135
FT 928 135
FT 929 135
FT 930 135
FT 931 135
FT 932 135
FT 933 135
FT 934 135
FT 935 135
FT 936 135
FT 937 135
FT 938 135
FT 939 135
FT 940 135
FT 941 135
FT 942 135
FT 943 135
FT 944 135
FT 945 135
FT 946 135
FT 947 135
FT 948 135
FT 949 135
FT 950 135
FT 951 135
FT 952 135
FT 953 135
FT 954 135
FT 955 135
FT 956 135
FT 957 135
FT 958 135
FT 959 135
FT 960 135
FT 961 135
FT 962 135
FT 963 135
FT 964 135
FT 965 135
FT 966 135
FT 967 135
FT 968 135
FT 969 135
FT 970 135
FT 971 135
FT 972 135
FT 973 135
FT 974 135
FT 975 135
FT 976 135
FT 977 135
FT 978 135
FT 979 135
FT 980 135
FT 981 135
FT 982 135
FT 983 135
FT 984 135
FT 985 135
FT 986 135
FT 987 135
FT 988 135
FT 989 135
FT 990 135
FT 991 135
FT 992 135
FT 993 135
FT 994 135
FT 995 135
FT 996 135
FT 997 135
FT 998 135
FT 999 135
FT 1000 135

```

FT DISULFID 129 145 BY SIMILARITY.
 FT DISULFID 147 156 BY SIMILARITY.
 FT DISULFID 163 175 BY SIMILARITY.
 FT DISULFID 169 184 BY SIMILARITY.
 FT DISULFID 186 195 BY SIMILARITY.
 FT DISULFID 202 213 BY SIMILARITY.
 FT DISULFID 207 223 BY SIMILARITY.
 FT DISULFID 225 234 BY SIMILARITY.
 FT DISULFID 241 252 BY SIMILARITY.
 FT DISULFID 246 261 BY SIMILARITY.
 FT DISULFID 263 272 BY SIMILARITY.
 FT DISULFID 279 292 BY SIMILARITY.
 FT DISULFID 286 301 BY SIMILARITY.
 FT DISULFID 303 312 BY SIMILARITY.
 FT DISULFID 319 330 BY SIMILARITY.
 FT DISULFID 324 339 BY SIMILARITY.
 FT DISULFID 341 350 BY SIMILARITY.
 FT DISULFID 356 367 BY SIMILARITY.
 FT DISULFID 361 378 BY SIMILARITY.
 FT DISULFID 380 389 BY SIMILARITY.
 FT DISULFID 396 409 BY SIMILARITY.
 FT DISULFID 403 418 BY SIMILARITY.
 FT DISULFID 420 429 BY SIMILARITY.
 FT DISULFID 436 447 BY SIMILARITY.
 FT DISULFID 441 456 BY SIMILARITY.
 FT DISULFID 458 467 BY SIMILARITY.
 FT DISULFID 474 485 BY SIMILARITY.
 FT DISULFID 479 494 BY SIMILARITY.
 FT DISULFID 496 505 BY SIMILARITY.
 FT DISULFID 512 523 BY SIMILARITY.
 FT DISULFID 517 532 BY SIMILARITY.
 FT DISULFID 534 543 BY SIMILARITY.
 FT DISULFID 550 560 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 571 580 BY SIMILARITY.
 FT DISULFID 587 598 BY SIMILARITY.
 FT DISULFID 592 607 BY SIMILARITY.
 FT DISULFID 609 618 BY SIMILARITY.
 FT DISULFID 625 635 BY SIMILARITY.
 FT DISULFID 630 644 BY SIMILARITY.
 FT DISULFID 646 655 BY SIMILARITY.
 FT DISULFID 662 673 BY SIMILARITY.
 FT DISULFID 667 682 BY SIMILARITY.
 FT DISULFID 684 693 BY SIMILARITY.
 FT DISULFID 700 710 BY SIMILARITY.
 FT DISULFID 705 719 BY SIMILARITY.
 FT DISULFID 721 730 BY SIMILARITY.
 FT DISULFID 739 750 BY SIMILARITY.
 FT DISULFID 744 759 BY SIMILARITY.
 FT DISULFID 761 770 BY SIMILARITY.
 FT DISULFID 776 787 BY SIMILARITY.
 FT DISULFID 781 797 BY SIMILARITY.
 FT DISULFID 799 808 BY SIMILARITY.
 FT DISULFID 815 827 BY SIMILARITY.
 FT DISULFID 821 836 BY SIMILARITY.
 FT DISULFID 838 847 BY SIMILARITY.
 FT DISULFID 854 865 BY SIMILARITY.
 FT DISULFID 859 874 BY SIMILARITY.
 FT DISULFID 876 885 BY SIMILARITY.
 FT DISULFID 892 902 BY SIMILARITY.
 FT DISULFID 897 911 BY SIMILARITY.
 FT DISULFID 913 922 BY SIMILARITY.
 FT DISULFID 929 940 BY SIMILARITY.
 FT DISULFID 934 949 BY SIMILARITY.
 FT DISULFID 951 960 BY SIMILARITY.
 FT DISULFID 967 978 BY SIMILARITY.
 FT DISULFID 972 987 BY SIMILARITY.
 FT DISULFID 989 998 BY SIMILARITY.
 FT DISULFID 1005 1016 BY SIMILARITY.
 FT DISULFID 1010 1023 BY SIMILARITY.
 FT DISULFID 1025 1034 BY SIMILARITY.
 FT DISULFID 1041 1062 BY SIMILARITY.
 FT DISULFID 1056 1071 BY SIMILARITY.

FT DISULFID 1073 1082 BY SIMILARITY.
 FT DISULFID 1089 1100 BY SIMILARITY.
 FT DISULFID 1094 1109 BY SIMILARITY.
 FT DISULFID 1111 1120 BY SIMILARITY.
 FT DISULFID 1127 1138 BY SIMILARITY.
 FT DISULFID 1132 1147 BY SIMILARITY.
 FT DISULFID 1149 1158 BY SIMILARITY.
 FT DISULFID 1165 1183 BY SIMILARITY.
 FT DISULFID 1177 1192 BY SIMILARITY.
 FT DISULFID 1194 1203 BY SIMILARITY.
 FT DISULFID 1210 1223 BY SIMILARITY.

Query Match 16.7%; Score 70; DB 1; Length 2318;
 Best Local Similarity 37.7%; Pred. No. 38;
 Matches 23; Conservative 2; Mismatches 24; Indels 12; Gaps 1;

QY 6 LPRPAVPPPLMOPGPAHPVLSFLRPSNDVLSARYSLPLAPLSTVPSISVSGRGP 65
 Db 2162 LNPVAVPLDMARLPPAPGPSFL-----LPLAPQULNPGAVSPQSRPP 2209
 QY 66 P 66
 Db 2210 P 2210

RESULT 15
 TEFS_CHICK STANDARD; PRT: 433 AA.
 ID TEFS_CHICK 090701; Q90702;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TRANSCRIPTIONAL ENHANCER FACTOR TEF-5 (CARDIAC-ENRICHED TEA DOMAIN
 DE TRANSCRIPTION FACTOR 1) (DREF-1).
 GN TEFS OR DTEF1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=96215227; PubMed=8626520;
 RA Azakie A., Larkin S.B., Farrance I.K., Greeningloh G., Ordahl C.P.;
 "DREF-1, a novel member of the transcription enhancer factor-1
 (TEF-1) multigene family.";
 RL J. Biol. Chem. 271:8260-8265(1996).
 CC -1- FUNCTION: SEQUENCE SPECIFIC M-CAT-BINDING FACTOR.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DREF-1A (SHOWN HERE) AND DTEF-
 CC 1B: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN CARDIAC MUSCLE, LOW IN SKELETAL
 CC MUSCLE. INTERMEDIATE LEVELS IN GIZZARD AND LUNG, LOW LEVELS IN
 CC KIDNEY.
 CC -1- SIMILARITY: CONTAINS A 'TEA' DNA-BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U46127; AAC59786.1; ALT_INIT.
 CC EMBL: U46128; AAC59787.1; ALT_INIT.
 CC InterPro: IPR000818; TEA.
 CC Pfam: PF01285; TEA.1.
 CC PRINTS: PR00065; TEADOMAIN.
 CC SMART: SM00426; TEA.1.
 CC PROSITE: PS00534; TEA_DOMAIN; 1.
 DR Transcription regulation; Activator; DNA-binding; Nuclear protein;
 KW

FW	Alternative splicing.	PRO-RICH (HYDROPHOBIC).
FT	DOMAIN 146 210	TRANSCRIPTIONAL ACTIVATION (POTENTIAL).
FT	DOMAIN 171 433	TEA-DOMAIN.
FT	DNA_BIND 30 97	POLY-PRO.
FT	DOMAIN 146 151	KROVSHIQLVLAIRISGDSKTK -> KROVSHIQLVLAIR
FT	VARSPLIC 87 110	KRYRSTGRHQ (IN ISOFORM DTEF-1B).
SO	SEQUENCE 433 AA; 48514 MW; 81077A9C4B0F245C CRC64;	

Query Match	16.6%	Score	69.5	DB	1	Length	433
Best Local Similarity	35.0%	Pred. No.	7.2				
Matches	21	Conservative	11	Mismatches	17	Gaps	4

QY 4 SLLEPPAAVNPVLPKMPGPAHVLSFLRPSMDLVS-----AFYSLPLAPLSP--TSVPI 55
| | : : | : : : : | : : : : | : : : :
Db 156 SAARFWSGPIP--GQPGPSODIKPFAOPAYPIQPPMPBSLASYE-PLAPLPAAASAVPV 212

Search completed: April 11, 2002, 09:34:18
Job time: 305 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:29:43 ; Search time 40.45 Seconds
(without alignments)
285.674 Million cell updates/sec

Title: US-09-234-208b-1
Perfect score: 418
Sequence: 1 GTHSLPPRAAVPPLRMQP.....VGRGPPDAHAVVNLRYEG 79

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406	97.1	419	4 Q9UK79	Q9UK79 homo sapien
2	85	20.3	327	11 Q99J66	Q99J66 mus musculu
3	84.5	20.2	434	5 Q9G0J5	Q9G0J5 theileria p
4	84	20.1	995	11 Q356I5	Q356I5 mus musculu
5	81.5	19.5	816	11 Q70474	Q70474 rattus norv
6	81	19.4	1006	10 Q82761	Q82761 arabidopsis
7	81	19.4	1006	10 Q82761	Q82761 arabidopsis
8	78.5	18.8	955	12 Q91MW0	Q91MW0 ceropithic
9	78	18.7	122	12 Q9YLRO	Q9YLRO hepatitis e
10	78	18.7	763	2 Q9XDH2	Q9XDH2 mycobacteri
11	78	18.7	2319	11 Q9R172	Q9R172 rattus norv
12	77.5	18.5	217	3 Q9P3I0	Q9P3I0 neurospora
13	77.5	18.5	487	5 Q9V197	Q9V197 dirosophila
14	77.5	18.5	1212	2 Q9LIC8	Q9LIC8 streptomyc
15	77	18.4	122	12 Q366I2	Q366I2 swine hepat
16	77	18.4	503	5 Q17585	Q17585 caenorhabdi
17	76.5	18.3	1110	5 Q09493	Q09493 caenorhabdi
18	76	18.2	309	3 Q9P6R1	Q9P6R1 schizosacch
19	76	18.2	356	10 Q9SAK1	Q9SAK1 arabidopsis

20	76	18.2	1044	4 Q9NXC6	Q9NXC6 homo sapien
21	75	17.9	189	12 Q08404	Q08404 soybean dwa
22	75	17.9	189	12 Q87034	Q87034 soybean dwa
23	75	17.9	295	2 Q9KX07	Q9KX07 streptomyc
24	75	17.9	1081	4 Q9H8F3	Q9H8F3 homo sapien
25	75	17.9	1952	3 Q9P6R1	Q9P6R1 neurospora
26	74.5	17.8	894	10 Q9SUS3	Q9SUS3 arabidopsis
27	74.5	17.8	894	10 Q9M6S0	Q9M6S0 arabidopsis
28	74	17.7	384	11 Q9D789	Q9D789 mus musculu
29	73.5	17.6	574	4 Q9BT08	Q9BT08 homo sapien
30	73.5	17.6	1822	4 Q9P1V7	Q9P1V7 homo sapien
31	73.5	17.6	2321	12 Q9DGT6	Q9DGT6 turkey harp
32	73	17.5	123	12 Q9WLK2	Q9WLK2 hepatitis e
33	73	17.5	639	4 Q9H9M1	Q9H9M1 homo sapien
34	73	17.5	1180	5 Q9VRM2	Q9VRM2 dirosophila
35	73	17.5	1520	4 Q9Y4E0	Q9Y4E0 homo sapien
36	73	17.5	1664	4 Q9BZES	Q9BZES homo sapien
37	73	17.5	3503	5 Q24292	Q24292 dirosophila
38	72.5	17.3	518	2 Q53738	Q53738 mycobacteri
39	72.5	17.3	715	2 Q9F9V7	Q9F9V7 mycobacteri
40	72.5	17.3	1410	5 Q9GRM9	Q9GRM9 leishmania
41	72.5	17.3	1709	4 Q15047	Q15047 homo sapien
42	72	17.2	122	12 Q71146	Q71146 hepatitis e
43	72	17.2	489	4 Q9B0C3	Q9B0C3 homo sapien
44	72	17.2	680	12 Q9C0X4	Q9C0X4 chayote mos
45	72	17.2	1914	13 Q91008	Q91008 gallus gall

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	419 AA.
Q9UK79	Q9UK79	Q9UK79		
AC	Q9UK79	Q9UK79		
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	HERSTATIN.			
GN	HER-2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99415951; PubMed=10485918;			
RA	Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;			
RT	"The HER-2/neu receptor tyrosine kinase gene encodes a secreted			
RT	autoinhibitor.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.;			
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF177761; AAD56009.2; -			
DR	InterPro: IPR000494; EGFR_L.			
DR	InterPro: IPR002174; Furin-Like.			
DR	Pfam: PF00757; Furin-Like; 1.			
DR	Pfam: PF01030; Recep_L_domain; 1.			
DR	SMART: SMO0261; FU; 1.			
SQ	SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CNC64;			
Query Match	97.1%; Score 406; DB 4; Length 419;			
Best Local Similarity	97.5%; Pred. No. 4.2e-34;			
Matches	77: Conservative 1; Mismatches 1; Indels 0; Gaps 0;			
OY	1 GTHSLPPRAAVPPLRMQPGAHVLSFLRSMQVLVAFSLPLAPSPISVPSV 60			
DB	341 GTHSLPPRAAVPPLRMQPGAHVLSFLRSMQVLVAFSLPLAPSPISVPSV 400			
OY	61 GRCPPDAHAVVNLRYEG 79			

Db 401 GRGPDPAHVAVDLSRYEG 419

RESULT 2
ID 099JK6 PRELIMINARY; PRT; 327 AA.
AC 099JK6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:3591061) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC006054; AAH06054.1; -
FT NON_TER 1
SQ SEQUENCE 327 AA; 33661 MW; 27917F16D583E774 CRC64;

Query Match 20.3%; Score 85; DB 11; Length 327;
Best Local Similarity 42.4%; Pred. No. 0.35;
Matches 28; Conservative 4; Mismatches 28; Indels 6; Gaps 4;

OY 1 GTHSLIPRAVPLRMQGPAPHVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPV 60
DB 88 GASGSCSP-PCRCPAPLRPODSPTNPAMSPRRPARGLDAA--SSP--PLEGSPSPSP-PA 141
OY 61 GRGPD 66
DB 142 GLSP 147

RESULT 3
ID 09GUIS PRELIMINARY; PRT; 434 AA.
AC 09GUIS;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 49.7 KDA PROTEIN.
OS Theileria parva.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC Theileria.
OX NCBI_TaxID=5875;
RN [1]
RP SEQUENCE FROM N.A.
RA Bishop R., Godlight E., Nene V., Morzaria S., Musoke A., Sohanpal B.;
RT "Polymorphic open reading frames encoding secretory proteins are located less than 3 kilobases from Theileria parva telomeres."
RL Mol. Biochem. Parasitol. 110:359-371(2000).
DR EMBL: AF225701; AAG28022.1; -
DR InterPro: IPR002965; P_rich_extensn.
DR PRINTS: PR01217; PRICHEXTENSN.
KW Hypothetical protein.
SQ SEQUENCE 434 AA; 49670 MW; 0F46586F4021A22F CRC64;

Query Match 20.2%; Score 84.5; DB 5; Length 434;
Best Local Similarity 31.0%; Pred. No. 0.53;
Matches 26; Conservative 11; Mismatches 34; Indels 13; Gaps 4;

OY 4 SLIPRAVPLV--PLRMQGP-PAHPVLSFLRPSMDLVSAF-----SLPLAPLSPTSPV 55
DB 159 ALXPRLPIOPPLPLQGHAPRP-----PRYPVSGTSPNHYAPPSVPQPIPPS 213
OY 56 SPVSGRGPDPDAHVAVDLSRYEG 79

Db 214 THVPTPTQPOPOQVPOQVPOQYGYG 237

RESULT 4
ID 035615 PRELIMINARY; PRT; 995 AA.
AC 035615;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FRIEND OF GATA-1 (FOG).
GN ZFPML OR FOG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373824; PubMed=9230307;
RA Tsang A.P., Visvader J.E., Turner C.A., Fujiwara Y., Yu C.,
RA Weiss M.J., Crossley M., Orkin S.H.;
RT "FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation."
RT Cell 90:109-119(1997).
RL -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC EMBL: AF006492; AAC53292.1; -
DR MGD: MGI:1095400; Zfpml.
DR InterPro: IPR001005; MYB_DNA_bind.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; zfc-C2H2; 9.
DR PRINTS: SM00355; ZNF_C2H2; 5.
DR SMART: SM00355; ZNF_C2H2; 5.
DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_2.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 2.
KW DNA-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 995 AA; 105983 MW; 293255B28151ECB8 CRC64;

Query Match 20.1%; Score 84; DB 11; Length 995;
Best Local Similarity 33.3%; Pred. No. 1.4;
Matches 23; Conservative 10; Mismatches 28; Indels 8; Gaps 2;

OY 7 PRPAVPLVPLRMQGPAPHVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVGRGP-- 64
DB 766 PPAPGAPVAVPVPSPTALFSPSPR-----GSASGAPAPALSPSPVDPGPIIDLSKRPR 820
OY 65 -DPDAHVA 72
DB 821 QSPDAPAL 829

RESULT 5
ID 070474 PRELIMINARY; PRT; 816 AA.
AC 070474;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NEUROCAN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HIPPOCAMPUS;
RA Zachmann-Brand B., Schaller H.C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF060879; AAC15766.1; -
DR HSSP: F00740; 1EDM.

DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR000538; Link.
 DR Pfam: PF00008; EGF_2.
 DR Pfam: PF00193; Xlink_1.
 DR PRINTS: PRO0010; EGFBLD.
 DR PRINTS: PRO1265; LINKMODULE.
 DR ProDom: PD00918; Link_1.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00001; EGF_like; 1.
 DR SMART: SM00445; Link; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE: PS50041; C_Type_Lectin_2; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_3.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS01241; Link; 1.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
 FT NON_TER 1
 FT 1
 SQ SEQUENCE 816 AA; 85578 MW; 957F5917AD10616E CRC64;

Query Match 19.5%; Score 81.5; DB 11; Length 816;
 Best Local Similarity 35.0%; Pred. No. 2.1;
 Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

QY 4 SLPPRAAVPVPLRMQPG---PAHPVLSFLR-----PSMDVSAFYSLPLAPLS--PT 51
 DB 342 SSTPSALSAVSIQSPGSGSPDFIVAMLRAPKMLPLHSTLVVSPPLSPASPPLPS 401
 QY 52 SVP---ISPVSGRGPPD 67
 DB 402 SYPEQAVRPVSFG-AEDPE 420

RESULT 6
 082761
 AC 082761 PRELIMINARY; PRT; 200 AA.
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE BLUE COPPER-BINDING PROTEIN.
 GN T16B12.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Rongning C.M., Fraser C.M., Somerville C.R., Venter J.C.,
 RT "Arabidopsis thaliana chromosome II BAC T16B12 genomic sequence."
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005311; AAC63847.1; -
 DR HSSP: P00303; 2CBP.
 DR InterPro: IPR003245; Cu_bind_like.
 DR Pfam: PF02298; Cu_bind_like; 1.
 DR ProDom: PD003122; Cu_bind_like; 1.
 SQ SEQUENCE 200 AA; 21475 MW; E669011C997E349C CRC64;

Query Match 19.4%; Score 81; DB 10; Length 200;
 Best Local Similarity 34.3%; Pred. No. 0.35;
 Matches 23; Conservative 5; Mismatches 23; Indels 16; Gaps 2;

QY 9 PAAPVPLRMQPGPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPISPVSGRGPDPDA 68
 DB 132 PYAAPV-----PGVAPRPSPSSPSS-----QSPLASPVNHAIVQRMGSPAP 175
 QY 69 HVAVNLS 75
 DB 176 HSNASNS 182

RESULT 7
 Q9LMQ1
 ID Q9LMQ1 PRELIMINARY; PRT; 1006 AA.
 AC Q9LMQ1
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE F7H2.17 PROTEIN.
 GN F7H2.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Liu S.X., Sakano H., Yu G., Etgu P., Lee J., Lanz C., Pham P.,
 RA Toriumi M., Chan A., Chung M., Goldsmith A., Liu A., Smith A.,
 RA Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
 RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
 RA Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
 RA Ecker J.R., Federspiel N.A., Theologis A., Theologis A., Theologis A.,
 RT "The sequence of BAC F7H2 from Arabidopsis thaliana chromosome 1."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC034256; AAF82153.1; -
 DR InterPro: IPR002965; P-rich-extension.
 DR PRINTS: PRO1217; PRICHEXTENS.
 SQ SEQUENCE 1006 AA; 103943 MW; C9FB49F9930C238D CRC64;

Query Match 19.4%; Score 81; DB 10; Length 1006;
 Best Local Similarity 31.4%; Pred. No. 2.9;
 Matches 22; Conservative 5; Mismatches 21; Indels 22; Gaps 2;

QY 6 LPPRAAVPVPLRMQPGPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPISPVSGRGPDP 65
 DB 336 LPPPATLPPPLPLPPPP-----SLPVPCSP---PPPIIVGAP 373
 QY 66 PDAHVAVNLS 75
 DB 374 PPCVTCVOVS 383

RESULT 8
 Q9IMY0
 ID Q9IMY0 PRELIMINARY; PRT; 955 AA.
 AC Q9IMY0
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE LATENT NUCLEAR ANTIGEN EBNA-3A.
 GN EBNA-3A.
 OS Cerepithicline herpesvirus 15.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=104228;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LC18664;
 RX MEDLINE=20304984; PubMed=10846073;
 RA Jiang H., Cho Y.-G., Wang F.,
 RT "Structural, Functional, and Genetic Comparisons of Epstein-Barr Virus
 Nuclear Antigen 3A, 3B, and 3C Homologs Encoded by the Rhesus

DR SMART; SM00248; ANK; 5

06 MCDL-1441-1441/
RN [1]

OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces
OX NCBI_TaxID=1902;
NN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Oliver K., Harris D.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RX MEDLINE-97000351; PubMed-8843436;

RA Redenbach M., Kieser H.M., Denaplatte D., Eichner A., Cullum J.,

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL157953; CAB/6073.1; -

DR InterPro; IPR003200; DBI_PRT.

DR InterPro; IPR000415; Nitroreductase.

DR InterPro; IPR002965; P_Rich_extensn.

DR Pfam; PF02277; DBI_PRT; 1.

DR Pfam; PF00881; Nitroreductase; 1.

DR PRINTS; PR01217; PRICHEXTENS.

SQ SEQUENCE 1212 AA; 124280 MW; E898B903F78D235A CRC64;

Query Match

Best Local Similarity 18.5%; Score 77.5; DB 2; Length 1212;

Matches 22; Conservative 7; Mismatches 28; Indels 3; Gaps 2;

OY 9 PAAVPLRMOPGPAHPVSLRPSWDLVSATFSLPLAPLSPTSVISPSVGRGPDPA 68
DB 488 PAAVPAVPAQEPHPAPQVGFPLVDGGVPT--THLAPTPPEAVLVPEETAAP-PEVYA 544

RESULT 15

036612

ID 036612; PRELIMINARY; PRT; 122 AA.

AC 036612; 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)

DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE HYPOTHETICAL 12.5 KDA PROTEIN (ORF-3).

OS Swine hepatitis E virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage.

OX NCBI_TaxID=63421;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-MENG;

RX MEDLINE-97420774; PubMed-9275216;

RA Meng X.J., Purcell R.H., Halbur P.G., Lehman J.R., Webb D.M.,

RA Tsareva T.S., Haynes J.S., Thacker B.J., Emerson S.U.;

RT "A novel virus in swine is closely related to the human hepatitis E

RT virus".

RL Proc. Natl. Acad. Sci. U.S.A. 94:9860-9865(1997).

DR EMBL; AF082843; AAC97209.1; -

DR InterPro; IPR003384; HEV_ORF2.

DR Pfam; PF02444; HEV_ORF2; 1.

SQ SEQUENCE 122 AA; 12466 MW; 77602F9048E7B12A CRC64;

Query Match

Best Local Similarity 18.4%; Score 77; DB 12; Length 122;

Matches 22; Conservative 10; Mismatches 23; Indels 6; Gaps 3;

OY 1 GTHTSLPRAAVPVPLRMOPGPAHPVSLRPSWDLVSATFSLPLAPL--SPTSVISPS 57
DB 55 GTTGLILSPS--PSPFIQPTPSLP-MSFHNGLERFALDSRAPLAPLGLVTSAPSAPPLPP 111

OY 58 V 58

DB 112 V 112

Search completed: April 11, 2002, 09:35:06
Job time: 323 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:28:23 ; Search time 20.27 Seconds
(without alignments)
87.704 Million cell updates/sec

Title: US-09-234-208B-1

Perfect score: 418

Sequence: 1 GTHSLPRPAAPVPLRMQP.....VGRGPDPAHVAVLSRYEG 79

Scoring table:

BLOSUM62

Gap 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/CTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfills1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81.5	19.5	1257	1 US-08-340-428B-49	Sequence 49, Appl
2	72	17.2	440	4 US-08-430-286A-9	Sequence 9, Appl
3	71.5	17.1	604	2 US-08-468-576B-12	Sequence 12, Appl
4	71.5	17.1	604	2 US-08-468-579B-12	Sequence 12, Appl
5	71.5	17.1	604	3 US-08-468-577B-12	Sequence 12, Appl
6	71.5	17.1	2441	1 US-08-194-468-2	Sequence 2, Appl
7	71.5	17.1	2441	3 US-08-961-739-2	Sequence 2, Appl
8	69.5	16.6	432	1 US-08-615-170-21	Sequence 21, Appl
9	69.5	16.6	433	1 US-08-615-170-19	Sequence 21, Appl
10	68.5	16.4	115	4 US-09-461-697-58	Sequence 58, Appl
11	68.5	16.4	132	4 US-09-461-697-52	Sequence 54, Appl
12	68.5	16.4	139	4 US-09-461-697-52	Sequence 52, Appl
13	68.5	16.4	139	4 US-09-461-697-48	Sequence 48, Appl
14	67.5	16.1	123	3 US-08-840-316-3	Sequence 3, Appl
15	67.5	16.1	123	3 US-08-478-507-9	Sequence 9, Appl
16	67.5	16.1	123	4 US-08-809-523-3	Sequence 3, Appl
17	67.5	16.1	123	4 US-09-128-275A-9	Sequence 9, Appl
18	67.5	16.1	123	4 US-08-471-971-3	Sequence 3, Appl
19	67.5	16.1	123	5 PCT-US93-08849A-3	Sequence 3, Appl
20	67.5	16.1	123	5 PCT-US93-08849-3	Sequence 3, Appl
21	67.5	16.1	124	4 US-08-240-049B-19	Sequence 19, Appl
22	67.5	16.1	124	4 US-08-542-634-21	Sequence 21, Appl
23	67.5	16.1	124	4 US-08-477-297-21	Sequence 21, Appl
24	67.5	16.1	124	5 PCT-US95-13703-21	Sequence 21, Appl
25	67.5	16.0	174	4 US-08-818-112-143	Sequence 143, Appl
26	66.5	15.9	1487	4 US-08-760-489-2	Sequence 2, Appl
27	66.5	15.9	1487	2 US-08-760-489-4	Sequence 4, Appl

28	66	15.8	325	4 US-09-041-886-33	Sequence 33, Appl
29	66	15.8	754	2 US-08-525-864A-2	Sequence 2, Appl
30	66	15.8	1447	4 US-09-041-886-25	Sequence 25, Appl
31	66	15.8	1447	5 PCT-US94-05277-2	Sequence 2, Appl
32	65.5	15.7	1257	4 US-09-220-641-3	Sequence 3, Appl
33	65	15.6	267	4 US-08-818-112-142	Sequence 142, App
34	65	15.6	1291	4 US-09-150-460B-10	Sequence 10, Appl
35	65	15.6	1291	4 US-09-220-641-5	Sequence 5, Appl
36	64.5	15.4	124	1 US-08-240-049B-20	Sequence 20, Appl
37	64.5	15.4	124	4 US-08-542-634-22	Sequence 22, Appl
38	64.5	15.4	124	4 US-08-477-292-22	Sequence 22, Appl
39	64.5	15.4	124	5 PCT-US95-13703-22	Sequence 22, Appl
40	64.5	15.4	543	2 US-08-469-412A-7	Sequence 7, Appl
41	64.5	15.4	543	4 US-09-021-715-7	Sequence 7, Appl
42	64.5	15.4	548	2 US-08-469-412A-2	Sequence 2, Appl
43	64.5	15.4	548	4 US-09-021-715-2	Sequence 2, Appl
44	64.5	15.4	1253	1 US-08-252-966B-12	Sequence 12, Appl
45	64.5	15.4	1261	1 US-08-252-966B-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648465
; GENERAL INFORMATION:
; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Dwe
; APPLICANT: MARGOLIS, Renee K.
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,428B
; FILING DATE: 14 No. 5648465ember 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/922,911
; FILING DATE: 03 August 1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: Margolis-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-340-428B-49
Query Match 19.5%; Score 81.5; DB 1; Length 1257;
Best Local Similarity 35.0%; Pred. No. 0.3;

Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

QY 4 SLIPRAAVPVPLRMQPG---PAHPVLSFLR-----PSMDVSAFYSLPLAPLS--PT 51
Db 610 SSISPSEALSAVSIQASPDGSPDPYVAMLRAPKMLPLPHSTLVPNVSPILPSAPSLPS 669

QY 52 SVP---ISPVSVGRGPPD 67
Db 670 SVPEEQAVRPVSFG-AEDPE 688

RESULT 2
US-08-430-286A-9
; Sequence 9, Application US/08430286A
; Patent No. 6225080
; GENERAL INFORMATION:
; APPLICANT: UHL, George R.
; APPLICANT: Eppler, C. Mark
; APPLICANT: Wang, Jai-Bel
; TITLE OF INVENTION: Mu-Subtype Opioid Receptor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dardy & Dardy PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,286A
; FILING DATE: 28-APR-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Robinson, Joseph R.
; REGISTRATION NUMBER: 33,448
; REFERENCE/DOCKET NUMBER: 0646/1A843-US5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 440 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: OPB-R
; US-08-430-286A-9

Query Match 17.2%; Score 72; DB 4; Length 440;
Best Local Similarity 45.2%; Pred. No. 1;
Matches 19; Conservative 5; Mismatches 10; Indels 8; Gaps 3;

QY 2 THSLPRAAVPVPLRMQPGPAHPVLSFLRPSMDVSAFYSL 43
Db 32 TASPSAPSWTSPSP---RPGPAHP---FLQPPMAV--ALMSL 65

RESULT 3
US-08-468-576B-12
; Sequence 12, Application US/08468576B
; Patent No. 5953345
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,576B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-576B-12

APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,576B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-468-576B-12

Query Match 17.1%; Score 71.5; DB 2; Length 604;
Best Local Similarity 34.8%; Pred. No. 1.8;
Matches 23; Conservative 13; Mismatches 27; Indels 3; Gaps 2;

QY 10 AAVPVPLRMQPGPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSVSPV---SVGRGPPD 67
Db 421 AAQPLNLSRKPATPEPKSPTPTONLFPAKTSPTVPLPKKSSIP-SPIGSLGRGSSLD 479

QY 68 AHYAVN 73
Db 480 ILSSLN 485

RESULT 4
US-08-468-579B-12
; Sequence 12, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,579B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-579B-12


```

1 ADDRESS: Sprung Kramer Schaefer & Briscoe
2 STREET: 660 White Plains Road
3 CITY: Tarrytown
4 STATE: New York
5 COUNTRY: USA
6 ZIP: 10591-5144
7
8 COMPUTER READABLE FORM:
9 MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
10 COMPUTER: Apple Macintosh
11 OPERATING SYSTEM: System 7.5
12 SOFTWARE: WordPerfect
13
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/468, 5798
16 FILING DATE: 06-JUN-1995
17
18 CLASSIFICATION: 530
19
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 08/239, 276
22 FILING DATE: 05-MAY-1994
23
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 07/872, 646
26 FILING DATE: 08-JUN-1992
27
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 07/715, 181
30 FILING DATE: 14-JUN-1991
31
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 07/441, 703
34 FILING DATE: 04-DEC-1989
35
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 07/312, 543
38 FILING DATE: 17-FEB-1989
39
40 ATTORNEY/AGENT INFORMATION:
41 NAME: Kurt G. Briscoe
42 REGISTRATION NUMBER: 33,141
43 REFERENCE/DOCKET NUMBER: JMOI 251.5-KGB
44
45 TELECOMMUNICATION INFORMATION:
46 TELEPHONE: (914) 332-1700
47 TELEFAX: (914) 332-1844
48
49 INFORMATION FOR SEQ ID NO: 12:
50
51 SEQUENCE CHARACTERISTICS:
52 LENGTH: 604 amino acids
53 TYPE: amino acid
54 TOPOLOGY: linear
55
56 US-08-468-579B-12
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940

```

```

      ZIP: 10591-5144
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
      COMPUTER: Apple Macintosh
      OPERATING SYSTEM: System 7.5
      SOFTWARE: Wordperfect
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/468,577B
      FILING DATE: 06-JUN-1995
      CLASSIFICATION: 514
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/239,276
      FILING DATE: 05-MAY-1994
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/872,646
      FILING DATE: 08-JUN-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/715,181
      FILING DATE: 14-JUN-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/441,703
      FILING DATE: 04-DEC-1989
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/312,543
      FILING DATE: 17-FEB-1989
      ATTORNEY/AGENT INFORMATION:
      NAME: Kurt G. Briscoe
      REGISTRATION NUMBER: 33,141
      REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (914) 332-1700
      TELEFAX: (914) 332-1844
      INFORMATION FOR SEO ID NO: 12:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 604 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      US-08-468-577B-12

Query Match          17.1% Score 71.5; DB 3; Length 604;
Best Local Similarity 34.8%; Pred. No. 1.8;
Matches 23; Conservative 13; Mismatches 27; Indels 3; Gaps 2;

QY    10 AAVPYLRMGCPHAPVLSFLRPSMDLVSAPYSLLAPLSPTSVSPV--SYGRCGDDP 67
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    421 AAQPLNTSSRRKTAEPVKSPTSPTQNLEPKASTSPVALPKKSSIP-SPIGSLGRSSLD 479

QY    68 AHVAVN 73
      ::|

Db    480 ILSSLIN 485

RESULT        6
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGENE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
STREET: Pretzy, Schroeder, Brueggemann & Clark
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/194,468
APPLICATION NUMBER: US/08/194,468
FILING DATE: 10-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9672
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)-546-4737
TELEFAX: (619)-546-9392
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2441 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-194-468-2

Query Match 17.1%; Score 71.5; DB 1; Length 2441;
Best Local Similarity 36.5%; Pred. No. 10;
Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

OY 4 SLPPRAAVPVLRMQPGAHVLSFLRPSMDLVSFAFSLPLAPLSPTSPVSVGSG 63
DB 843 SGLPCPPVYQSLHPTPPASTAGM--PSLQHPYAPGMPYPPQAPAPVQ-PSIPVSSGQT 899
OY 64 PDP 66
DB 900 PTP 902

RESULT 7
US-08-961-739-2
Sequence 2, Application US/08961739A
Patent No. 6063583
GENERAL INFORMATION:
APPLICANT: Montminy, Marc R.
TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
FILE REFERENCE: SALK1650-1
CURRENT APPLICATION NUMBER: US/08/961,739A
CURRENT FILING DATE: 1997-10-31
EARLIER APPLICATION NUMBER: US 194,468
EARLIER FILING DATE: 1994-02-10
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 2441
TYPE: PRT
ORGANISM: Mus
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(2441)
OTHER INFORMATION: Xaa = Any Amino Acid
US-08-961-739-2

Query Match 17.1%; Score 71.5; DB 3; Length 2441;
Best Local Similarity 36.5%; Pred. No. 10;
Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

OY 4 SLPPRAAVPVLRMQPGAHVLSFLRPSMDLVSFAFSLPLAPLSPTSPVSVGSG 63
DB 843 SGLPCPPVYQSLHPTPPASTAGM--PSLQHPYAPGMPYPPQAPAPVQ-PSIPVSSGQT 899
OY 64 PDP 66
DB 900 PTP 902

RESULT 8
US-08-615-170-21
Sequence 21, Application US/08615170
Patent No. 5776776
GENERAL INFORMATION:
APPLICANT: ORDAHL, Charles P.
APPLICANT: AZARIE, Anthony
APPLICANT: MAR, Janet H.
APPLICANT: FARRANCE, Iain K.G.
APPLICANT: HALL, Deborah E.
APPLICANT: STEWART, Alexandre F.R.
TITLE OF INVENTION: DTEF-1 ISOFORMS AND USES THEREOF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/615,170
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01526
FILING DATE: 06-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/191,493
FILING DATE: 04-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 23070-053120
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-615-170-21

Query Match 16.6%; Score 69.5; DB 1; Length 432;
Best Local Similarity 35.0%; Pred. No. 2;
Matches 21; Conservative 11; Mismatches 17; Indels 11; Gaps 4;

OY 4 SLPPRAAVPVLRMQPGAHVLSFLRPSMDLVS-----AFYSLPLAPLSP--TSVPI 55
DB 155 SNAPEWSPPIP--GQPGSDIKRFPQAPVPIQPPMPPLASYS-PLAPLPPASAVPV 211
RESULT 9
US-08-615-170-19
Sequence 19, Application US/08615170
Patent No. 5776776
GENERAL INFORMATION:
APPLICANT: ORDAHL, Charles P.
APPLICANT: AZARIE, Anthony
APPLICANT: MAR, Janet H.
APPLICANT: FARRANCE, Iain K.G.

```
; APPLICANT: HALL, Deborah E.
; APPLICANT: STEWART, Alexandre F.R.
; APPLICANT: LARKIN, Sarah B.
; TITLE OF INVENTION: DREF-1 ISOFORMS AND USES THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend Kourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,170
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01526
; FILING DATE: 06-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/191,493
; FILING DATE: 04-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 23070-053120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-615-170-19

Query Match          16.6%; Score 69.5; DB 1; Length 433;
Best Local Similarity 35.0%; Pred. No. 2;
Matches 21; Conservative 11; Mismatches 17; Indels 11; Gaps 4;

OY 4 SLRPPRAVPPLRMOPPAHVLSFLRPSMDVLS-----AFYSLPLAPLSP--TSVPI 55
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 156 SAAPFWGSPIP--GQPPSQDIKFPQAPVPIOPMPPLSLAYE-PLAPLPAPASAVPV 212

RESULT 10
US-09-461-697-58
; Sequence 58, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
```

```
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-461-697-58

Query Match          16.4%; Score 68.5; DB 4; Length 115;
Best Local Similarity 27.8%; Pred. No. 0.5;
Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

OY 1 GTHSLPPRAVPVPLRMOP-----GPAH--PVLSFLRPSMDL-----VSAFY 41
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 11 GCHCMMSRDLTPLPHPSEPGVLDCLGFCCHLLPLSPSPCWCWGLHSLRPPSAASH 70

OY 42 SLPLAPLSPTSVISPVSVS-----VGRG 63
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 71 ALTITSLPPGLLPFGVGLTAHPQALLIGRG 100

RESULT 11
US-09-461-697-54
; Sequence 54, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-461-697-54

Query Match          16.4%; Score 68.5; DB 4; Length 132;
Best Local Similarity 27.8%; Pred. No. 0.59;
Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

OY 1 GTHSLPPRAVPVPLRMOP-----GPAH--PVLSFLRPSMDL-----VSAFY 41
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 28 GCHCMMSRDLTPLPHPSEPGVLDCLGFCCHLLPLSPSPCWCWGLHSLRPPSAASH 87

OY 42 SLPLAPLSPTSVISPVSVS-----VGRG 63
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 88 ALTITSLPPGLLPFGVGLTAHPQALLIGRG 117

RESULT 12
US-09-461-697-52
; Sequence 52, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
```

APPLICANT: Katz, Lawrence C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
TITLE OF INVENTION: CELL DEATH
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461,697
CURRENT FILING DATE: 1999-12-14
NUMBER OF SEQ ID NOS: 466
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 52
LENGTH: 139
TYPE: PRT
ORGANISM: Homo sapiens
US-09-461-697-52

Query Match 16.4%; Score 68.5; DB 4; Length 139;
Best Local Similarity 27.8%; Pred. No. 0.63;
Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

OY 1 GTHSLPRAAVPVLRLMOP-----GPAH--PVLSFLRPSMDL-----VSARY 41
DB 35 GCHCMSRDLTPPLPSPSPGVLDCLGPHLLPLSPGSPCVWLGLHFSLHPPSAASH 94
OY 42 SLPLAPLSPVSPISPV-----VGRG 63
DB 95 ALTITSLPPGLLPFVGVELTAHPQALIGRG 124

RESULT 13
US-09-461-697-48
Sequence 48, Application US/09461697
Patent No. 6277974
GENERAL INFORMATION:

APPLICANT: COGENT NEUROSCIENCE, Inc.
APPLICANT: LO, Donald C.
APPLICANT: Barney, Shawn
APPLICANT: Thomas, Mary Beth
APPLICANT: Portbury, Stuart D.
APPLICANT: Puranam, Kasturi
APPLICANT: Katz, Lawrence C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461,697
CURRENT FILING DATE: 1999-12-14
NUMBER OF SEQ ID NOS: 466
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 48
LENGTH: 159
TYPE: PRT
ORGANISM: Homo sapiens
US-09-461-697-48

Query Match 16.4%; Score 68.5; DB 4; Length 159;
Best Local Similarity 27.8%; Pred. No. 0.74;
Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

OY 1 GTHSLPRAAVPVLRLMOP-----GPAH--PVLSFLRPSMDL-----VSARY 41
DB 55 GCHCMSRDLTPPLPSPSPGVLDCLGPHLLPLSPGSPCVWLGLHFSLHPPSAASH 114
OY 42 SLPLAPLSPVSPISPV-----VGRG 63
DB 115 ALTITSLPPGLLPFVGVELTAHPQALIGRG 144

RESULT 14
US-08-840-316-3
Sequence 3, Application US/08840316
Patent No. 6054567

GENERAL INFORMATION:

APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
APPLICANT: Tsarev, Sergei, A., and Robinson, Robin A.
TITLE OF INVENTION: Recombinant Proteins of
TITLE OF INVENTION: A Pakistani Strain of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,316
FILING DATE: 11-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Richard W. Bork
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4255
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acid residues
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
US-08-840-316-3

Query Match 16.1%; Score 67.5; DB 3; Length 123;
Best Local Similarity 34.3%; Pred. No. 0.71;
Matches 23; Conservative 8; Mismatches 17; Indels 19; Gaps 3;

OY 10 AAVPV-----PLRMQGPAAHVLFLRPSMDLVA--FYSPLAPLSPT 51
DB 48 AAVPAVSGVTGLILSPSQPIFIQPTPS-PMSPDLRGLDVLVFNPPDHSAPLGVTRPS 106
OY 52 SVPSISPV 58
DB 107 APPLPHV 113

RESULT 15
US-08-478-507-9
Sequence 9, Application US/08478507
Patent No. 6120988
GENERAL INFORMATION:

APPLICANT: Reyes, Gregory R
APPLICANT: Yarbough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk E
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250

```

1      CITY: Palo Alto
2      STATE: CA
3      COUNTRY: USA
4      ZIP: 94306
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE: Floppy disk
8      COMPUTER: IBM PC compatible
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: Patent Release #1.0, Version #1.25
11
12     CURRENT APPLICATION DATA:
13     . APPLICATION NUMBER: US/08/478,507
14     . FILING DATE: 07-JUN-1995
15     . PRIOR APPLICATION DATA:
16     . APPLICATION NUMBER: US 08/279,823
17     . FILING DATE: 25-JUL-1994
18     . PRIOR APPLICATION DATA:
19     . APPLICATION NUMBER: US 07/681,078
20     . FILING DATE: 05-APR-1991
21     . PRIOR APPLICATION DATA:
22     . APPLICATION NUMBER: US 07/505,888
23     . FILING DATE: 05-APR-1990
24     . PRIOR APPLICATION DATA:
25     . APPLICATION NUMBER: US 07/420,921
26     . FILING DATE: 13-OCT-1989
27     . PRIOR APPLICATION DATA:
28     . APPLICATION NUMBER: US 07/367,486
29     . FILING DATE: 16-JUN-1989
30     . PRIOR APPLICATION DATA:
31     . APPLICATION NUMBER: US 07/336,672
32     . FILING DATE: 11-APR-1989
33     . PRIOR APPLICATION DATA:
34     . APPLICATION NUMBER: US 07/208,997
35     . FILING DATE: 17-JUN-1988
36     . ATTORNEY/AGENT INFORMATION:
37     . NAME: Sholtz, Charles K.
38     . REGISTRATION NUMBER: 38,615
39     . REFERENCE/DOCKET NUMBER: 4600-0183.22
40     . TELECOMMUNICATION INFORMATION:
41     . TELEPHONE: (650) 324-0880
42     . TELEFAX: (650) 324-0960
43     . INFORMATION FOR SEO ID NO: 9:
44     . SEQUENCE CHARACTERISTICS:
45     . LENGTH: 123 amino acids
46     . TYPE: amino acid
47     . TOPOLOGY: Linear
48     . MOLECULE TYPE: protein
49
50     US-08-478-507-9

```

	Query Match	16.1%	Score 67.5;	DB 3;	Length 123;	
	Best Local Similarity	34.3%;	Pred. No. 0.71;			
	Matches	23; Conservative	8; Mismatches	17; Indels	19; Gaps	3;
Oy	10 AAVPV-----PLRMQGPANPRLSLRPSWDLVSA---EYSLPLANLSPT	51				
		::: :	::			::
Dd	48 AAAPVVGSGVTGLIISPGQSPIFIQTPTS-PPMSPLRLGDLVFANRPDHSAPLGVTRPS	106				
Oy	52 SVPISPV	58				
	: ::					
Dd	107 AAPPLPHV	113				

Search completed: April 11, 2002, 09:29:35
Job time: 72 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:29:08 ; Search time 40.19 seconds
(without alignments)
772.250 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287
Sequence: 1 MELALCRRGGLLALLPGEA.....VGRGPDPAHVAVNLSRYEG 419

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2158.5	94.4	420	AA97240	Truncated HER-2, p
2	1878	82.1	645	AA860408	Human ErbB2 oncopr
3	1878	82.1	645	AA861593	Human ErbB2 extrac
4	1878	82.1	653	AA821200	Extracellular HER-
5	1878	82.1	712	AA821204	Human HER-2/neu fu
6	1878	82.1	782	AA819764	HER-2/neu protein
7	1878	82.1	919	AA821203	Human HER-2/neu fu
8	1878	82.1	1200	AA821208	Human HER-2/neu fu
9	1878	82.1	1255	AA801111	HER-2/neu protein
10	1878	82.1	1255	AA92406	Human HER-2/neu on
11	1878	82.1	1255	AA821198	Human HER-2/neu pr

12	1878	82.1	1255	AA84780	Amino acid sequenc
13	1878	82.1	1255	AA92620	Human heregulin 2
14	1878	82.1	1255	AA85458	Human HER-2/neu pr
15	1878	82.1	1255	AA888267	HER2/neu amino aci
16	1878	82.1	1255	AA860167	HER2 transgene pla
17	1860	81.3	1433	AA839568	Sequence of c-erbB
18	1775	77.6	951	AA44993	DC8scfV-erbB2EC fu
19	1736	75.9	624	AA808222	Extracellular port
20	1608.5	70.3	654	AA821205	Rat HER-2/neu prot
21	1608.5	70.3	1256	AA821199	Rat HER-2/neu prot
22	1597.5	69.9	1256	AA821206	Mouse Her-2/neu pr
23	1597.5	69.9	1256	AA862860	Amino acid sequenc
24	987	43.2	191	AA859354	Human ErbB2 protei
25	987	43.2	191	AA41572	Human ErbB2 recept
26	987	43.2	191	AA848767	Human ErbB2 N-term
27	860	37.6	166	AA859345	Human ErbB2 domain
28	860	37.6	166	AA862074	ErbB2 protein doma
29	860	37.6	166	AA848763	Human ErbB2 extrac
30	793	34.7	405	AA833737	Epidermal growth f
31	793	34.7	529	AA838209	Epidermal growth f
32	793	34.7	1210	AA819259	LD1D2D3.Appl EGF r
33	793	34.7	1210	AA850616	Amino acid sequenc
34	793	34.7	1210	AA868420	Amino acid sequenc
35	782	34.2	621	AA867309	Epidermal growth f
36	775	33.9	910	AA891737	HER4-Ig fusion pro
37	775	33.9	1058	AA854843	HER4 with alternat
38	775	33.9	1058	AA891734	Receptor tyrosine
39	775	33.9	1292	AA802425	Human epidermal gr
40	775	33.9	1298	AA802423	Human epidermal gr
41	775	33.9	1308	AA854841	HER4. Homo sapien
42	775	33.9	1308	AA891733	Receptor tyrosine
43	775	33.9	1308	AA864499	Human inducible ni
44	773	33.8	478	AA848033	EGF receptor relat
45	735.5	32.2	1342	AA88453	erbB-3 polypeptide

ALIGNMENTS

RESULT 1	
AA97240	AA97240 standard; protein; 420 AA.
XX	
AC	AA97240:
XX	
DT	04-DEC-2000 (first entry)
XX	
DE	Truncated HER-2, p68-HER-2.
XX	
KW	HER-2, erbB-2; oncogene; receptor-like tyrosine kinase; insertion;
KW	extracellular domain IIIa; antagonist; intron 8; C-terminal extension;
KW	truncated HER-2; p68; dimerization inhibitor; cytosolic.
XX	
OS	Homo sapiens.
XX	
XX	
FT	Key
FT	Misc-difference 124
FT	/note= "represented as Agn"
FT	Misc-difference 125
FT	/note= "represented as Agn"
FT	Misc-difference 343
FT	/note= "Preferably Ser"
FT	Misc-difference 346
FT	/note= "Preferably Pro"
FT	Misc-difference 347
FT	/note= "Preferably Leu"
FT	Misc-difference 357
FT	/note= "Preferably Gln"
FT	Misc-difference 359
FT	/note= "Preferably Leu"
FT	Misc-difference 362
FT	/note= "changes from glycine"
FT	Misc-difference 377

[illegible]

Db	241	aagctgpkhscclacclhnhsgicelhocpalvlyntdfescmcppegyitfgascvtacp	300
Qy	301	YNYISTDVGSCTLVCPLHNOEVTAEADGTORCEKCSKPCAR-GTHSLPRAPAVPPPLRMQ	359
Db	301	Ynk1stdvgsctclvcplhngvetaedgfgicekcskpcarvaxhsxxprrpaavpvrpxrxq	360
Qy	360	PGPAHPVLSFLRPSMDIYSAFYSLPLAPLSPVSPISVSGRGDDPPAHVAVNLSPREG	419
Db	361	pxpahpvlstflrpswxdvsaftyslplaplclpsvxispsvsgrgddpahnvaavxlstrey	420
RESULT	2		
ID	AA060408		
XX	AA060408	standard; Protein: 645 AA.	
XX	AA060408;		
XX	24-APR-2001	(first entry)	
DE	Human ErbB2 oncoprotein, SEQ ID NO:13.		
XX	Anti-ErbB2 monoclonal antibody 2C4; HER2; mouse; murine; humanised; VL;		
XX	light chain variable region; cancer; cytosolic; EGFR-expressing cancer;		
KW	epidermal growth factor receptor; colon cancer; rectal cancer; tumour;		
KW	colorectal cancer; non-small cell lung cancer; metastatic breast cancer;		
KW	affinity purification.		
OS	Homo sapiens.		
XX	WO200100245-A2.		
PN	04-JAN-2001.		
PD	23-JUN-2000; 2000WO-US17366.		
XX	25-JUN-1999; 99US-0141316.		
PR	(GETH) GENENTECH INC.		
XX	Adams CW, Presta LG, Sliwkowsky M;		
PA	WPI, 2001-080862/09.		
PI	Treating cancer in a human, where the cancer expresses epidermal growth		
XX	factor receptor (EGFR), comprises administering an antibody which binds		
XX	ErbB2 -		
XX	Example 1; Fig 1A; 89pp; English.		
XX	The invention relates to a method for treating cancer in a human patient,		
CC	wherein the cancer expresses epidermal growth factor receptor (EGFR),		
CC	comprising administering an antibody which binds ErbB2 (HER2; AA060408).		
CC	In particular, the anti-ErbB2 antibody is the murine monoclonal antibody		
CC	2C4 (AA060396, AA060397) or a humanised version of 2C4 (AA060398,		
CC	AA060399). The invention also encompasses an isolated nucleic acid		
CC	encoding a humanised ErbB2-binding antibody; vectors and host cells		
CC	comprising such nucleic acids; the recombinant production of a humanised		
CC	ErbB2-binding antibody; and an immunocjugate comprising a humanised		
CC	ErbB2-binding antibody and a cytotoxic drug. The ErbB2-binding antibodies		
CC	act by antagonising ErbB receptors, and as inhibitors of transforming		
CC	growth factor alpha (TGF-alpha)-activated mitogen activated protein		
CC	kinase (MAPK). The method of the invention is used for treating cancer,		
CC	especially colon cancer, rectal cancer, colorectal cancer, lung cancer		
CC	(especially non-small cell lung cancer), or breast cancer (especially		
CC	metastatic breast cancer). The antibodies may also have non-therapeutic		
CC	uses e.g., as affinity purification agents. Using an antibody which binds		
CC	to ErbB2 to treat cancer is preferably to the use of EGFR-targeted		
CC	drugs, as EGFR is also highly expressed in other tissues such as the		
CC	liver and skin, where the active drug will also bind, with skin toxicity		
CC	having been observed for EGFR-targeted drugs. Antibodies which bind		
CC	ErbB2 are anticipated to have a better safety profile than such drugs.		
XX	The present sequence represents human ErbB2.		

SQ Sequence 645 AA:
 Query Match 82.1%; Score 1878; DB 22; Length 645;
 Best Local Similarity 83.0%; Pred. No. 1.7e-143;
 Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
 1 MELALCRWGLLALLPRGASTQVCTGTDMLRIPASPEHLDMLRHLYGCGVVOGNTL 60
 1 melalcrwgl11allppgaastqvcitgtdmklripaspehldmlrhllygqcvvgvgnl 60
 61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVROVPLQRLRIVRGTOLEEDNYALAVLDNG 120
 61 eltylptnaslsflqdiqevgyvliahnqvrvplqlrlrivrgtqlfednyalavldng 120
 121 DPLNNTPTVTGASPGGLRELDLRSLTEILKGVLIOQNPOLCYODTILMKDIFHKNNOLA 180
 121 dplnntptvtgaspzglrelqlrsltelikgvl1oqnpqlcyodtllwkdifhknq1a 180
 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGCARCKGPLPTDCCHQC 240
 181 ltlidtnsrachpcspmksgrcwgessecdqsltrtcaggarckgplptdcchqc 240
 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNEGRYTFGASCVTACP 300
 241 aagctgpkhsdclaclfhnsigicelhcpalvtynldfemnpnegrytfgascvtacp 300
 301 YNYLSTDVSGCTLVCPLEHNOEVTADGTQRCCKSKPCAR-----GTHSLPRPAAVPVP 355
 301 ynylstdvsgctlvcp1ehnoevtaedgtqrcckskpcarvcyglgmehlrevravtsan 360
 361 IRMORG--PAHPVLSFLRPMNDVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
 361 irfagcgckkifgslaf1pessifogdpasnt---ap1qpeqlqvfeleeltylyisawpd 417
 406 --PDAHVAVNLSRYEG 419
 418 slpdlsvfqn1qv1rg 433
 RESULT 3
 AAB61593
 ID AAB61593 standard; protein; 645 AA.
 AC AAB61593;
 DT 04-APR-2001 (first entry)
 DE Human ErbB2 extracellular domain.
 KW Human; ErbB2; cytosstatic; prostate cancer; receptor tyrosine kinase; antibody; ErbB receptor; monoclonal antibody 2C4; variable light chain.
 OS Homo sapiens.
 PN W0200100238-A1.
 PD 04-JAN-2001.
 PE 23-JUN-2000; 2000OWO-US17423.
 PR 25-JUN-1999; 99US-0141315.
 PA (GETH.) GENENTECH INC.
 PA (SLOK.) SLOAN KETTERING INST CANCER RES.
 PI Agus DB, Scher HI, Sliwkowski MX;
 WPI; 2001-159131/16.
 Treating prostate cancer in a human comprises administering an antibody which binds ErbB2 and blocks ligand activation of an ErbB receptor -

PS Disclosure; Fig 1; 93pp; English.
 CC The ErbB family of receptor tyrosine kinases are important mediators of cell growth, differentiation and survival. The receptor family includes four distinct members including Epidermal Growth Factor Receptor (EGFR or ErbB1), HER2 (ErbB2 or p185^{neu}), HER3 (ErbB3) and Her4 (ErbB4 or Tyro2). The present invention relates to a method for treating prostate cancer. The method comprises administering an antibody which binds ErbB2 and blocks ligand activation of an ErbB receptor. Preferably, the antibody blocks binding of monoclonal antibody 2C4 to ErbB2 and/or blocks TGF-alpha activation of mitogen-activated protein kinase (MAPK). The present sequence is the extracellular domain of human ErbB2.
 SQ Sequence 645 AA:
 Query Match 82.1%; Score 1878; DB 22; Length 645;
 Best Local Similarity 83.0%; Pred. No. 1.7e-143;
 Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
 1 MELALCRWGLLALLPRGASTQVCTGTDMLRIPASPEHLDMLRHLYGCGVVOGNTL 60
 1 melalcrwgl11allppgaastqvcitgtdmklripaspehldmlrhllygqcvvgvgnl 60
 61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVROVPLQRLRIVRGTOLEEDNYALAVLDNG 120
 61 eltylptnaslsflqdiqevgyvliahnqvrvplqlrlrivrgtqlfednyalavldng 120
 121 DPLNNTPTVTGASPGGLRELDLRSLTEILKGVLIOQNPOLCYODTILMKDIFHKNNOLA 180
 121 dplnntptvtgaspzglrelqlrsltelikgvl1oqnpqlcyodtllwkdifhknq1a 180
 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGCARCKGPLPTDCCHQC 240
 181 ltlidtnsrachpcspmksgrcwgessecdqsltrtcaggarckgplptdcchqc 240
 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNEGRYTFGASCVTACP 300
 241 aagctgpkhsdclaclfhnsigicelhcpalvtynldfemnpnegrytfgascvtacp 300
 301 YNYLSTDVSGCTLVCPLEHNOEVTADGTQRCCKSKPCAR-----GTHSLPRPAAVPVP 355
 301 ynylstdvsgctlvcp1ehnoevtaedgtqrcckskpcarvcyglgmehlrevravtsan 360
 361 IRMORG--PAHPVLSFLRPMNDVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
 361 irfagcgckkifgslaf1pessifogdpasnt---ap1qpeqlqvfeleeltylyisawpd 417
 406 --PDAHVAVNLSRYEG 419
 418 slpdlsvfqn1qv1rg 433
 RESULT 4
 AAB21200
 ID AAB21200 standard; protein; 653 AA.
 AC AAB21200;
 DT 12-JAN-2001 (first entry)
 DE Extracellular HER-2/neu protein.
 KW HER-2/neu; oncogene; tyrosine kinase; cytosstatic; vaccine; breast cancer; prostate cancer; ovarian cancer; lung cancer; colon cancer.
 OS unidentified.
 PN W0200044899-A1.
 PD 03-AUG-2000.

PF 28-JAN-2000; 2000WO-US02164.
XX
XX 29-JAN-1999; 99US-0117976.
XX
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHLINE BEECHAM.
XX
XX Cheever MA, Gheysen D;
XX
XX WPI: 2000-505976/45.
XX
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
XX useful for vaccinating against breast, ovarian, colon, lung and
XX prostate cancers -
XX
XX
XX Claim 2: Fig 9; 128pp; English.
XX
XX The present sequence is the extracellular HER-2/neu protein. HER-2/neu is
XX a member of the tyrosine kinase family of receptor-like glycoproteins and
XX shows homology to the epidermal growth factor receptor (EGFR). It
XX probably plays a part in cell growth and/or differentiation. The
XX HER-2/neu gene is an oncogene. An HER-2/neu fusion protein comprising a
XX HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
XX domain may be used to treat or prevent cancer by eliciting or
XX enhancing an immune response to the HER-2/neu protein. It may be used
XX to treat malignancies such as breast, ovarian, colon, lung and
XX prostate cancers, and may be used as an antigen to vaccinate against
XX these neoplasias.
XX
XX Sequence 653 AA;
SQ

Query Match 82.1%; Score 1878; DB 21; Length 653;
Best Local Similarity 83.0%; Pred. No. 1.7e-143;

Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRWGLLALLPFGAASVQCTGDMKRLRPASPEHLDMLRHLRYGCGVVGNL 60
DB 1 melaalcrwgl1allppgaastqctgdmkrlrpaspehldmlrhlrygscqvvgnl 60
QY 61 ELTYPTNASLSFLDIOVGQVYLIAHNOVQVPLQRLRIYRGVOLFEDNALAVLNG 120
DB 61 eltyptnaslsflldiovgqvylianhovqvpqlqrlriyrgvolfednyalavldng 120
QY 121 DLNMTPTVPGASPGELRELQRLSLTEILKGVLIQRLPOLCYODTIIMKDFHKNOLA 180
DB 121 dlnmtptvpgaspgelrelqrlslteilkgvliqrlpolcycodtiimkdfhknola 180
QY 121 dlnmtptvpgaspgelrelqrlslteilkgvliqrlpolcycodtiimkdfhknola 180
QY 181 LTLIDTNSRACHPCSPMKGSRGWESSEDCQSLRTVACAGCARCKGPLPTDCHEQC 240
DB 181 ltlidtnsrachpcspmkgsrgwessedcqsllrtvcagcarckgplptdccheqc 240
QY 241 AAGCTGPKHSDDLACLFHNSGICELHCPALVTYNTDFESMPNEGRITFGASCYTACP 300
DB 241 aagctgpkhsddlacLFHNSGICELHCPALVTYNTDFESMPNEGRITFGASCYTACP 300
QY 241 aagctgpkhsddlacLFHNSGICELHCPALVTYNTDFESMPNEGRITFGASCYTACP 300
QY 301 YVYLSIDVGSCTLVCPPLHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPAAVPVP 355
DB 301 yvylsidvgscTLVCPPLHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPAAVPVP 355
QY 301 yvylsidvgscTLVCPPLHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPAAVPVP 355
QY 356 LRMOFG--PAHFVLSFLRSPMDLVSAFYSLPAPLSPTVPI-----SPVSVGRGPD 405
DB 356 lrmoFG--PAHFVLSFLRSPMDLVSAFYSLPAPLSPTVPI-----SPVSVGRGPD 405
QY 361 lgefagckllfslafllpsfifgdpasnt---apldqeqldyfelteeltgyllylsawpd 417
DB 361 lgefagckllfslafllpsfifgdpasnt---apldqeqldyfelteeltgyllylsawpd 417
QY 406 ---PDAHVAVNLRYREG 419
DB 406 ---PDAHVAVNLRYREG 419
QY 418 sipdlsvfqnlgvixg 433
DB 418 sipdlsvfqnlgvixg 433

RESULT 5
AAB21204
ID AAB21204 standard; protein; 712 AA.
XX

AC AAB21204;
XX
XX 12-JAN-2001 (first entry)
XX
XX Human HER-2/neu fusion protein.
XX
XX Human HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
XX breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX colon cancer; fusion protein.
XX
XX Homo sapiens.
XX Synthetic.
XX WO200044899-A1.
XX
XX
XX 03-AUG-2000.
XX
XX 28-JAN-2000; 2000WO-US02164.
XX
XX 29-JAN-1999; 99US-0117976.
XX
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHLINE BEECHAM.
XX
XX Cheever MA, Gheysen D;
XX
XX WPI: 2000-505976/45.
XX
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
XX useful for vaccinating against breast, ovarian, colon, lung and
XX prostate cancers -
XX
XX
XX Claim 27: Fig 13; 128pp; English.
XX
XX The present sequence is a fusion protein comprising the extracellular
XX domain and a preferred portion of the phosphorylation domain of the human
XX HER-2/neu protein. HER-2/neu is a member of the tyrosine kinase family of
XX receptor-like glycoproteins and shows homology to the epidermal growth
XX factor receptor (EGFR). It probably plays a part in cell growth and/or
XX differentiation. The HER-2/neu gene is an oncogene. HER-2/neu fusion
XX proteins may be used to treat or prevent cancer by eliciting or enhancing
XX an immune response to the HER-2/neu protein. They may be used to treat
XX malignancies such as breast, ovarian, colon, lung and prostate cancers,
XX and may be used as an antigen to vaccinate against these neoplasias.
XX
XX Sequence 712 AA;
SQ

Query Match 82.1%; Score 1878; DB 21; Length 712;
Best Local Similarity 83.0%; Pred. No. 1.9e-143;

Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRWGLLALLPFGAASVQCTGDMKRLRPASPEHLDMLRHLRYGCGVVGNL 60
DB 1 melaalcrwgl1allppgaastqctgdmkrlrpaspehldmlrhlrygscqvvgnl 60
QY 61 ELTYPTNASLSFLDIOVGQVYLIAHNOVQVPLQRLRIYRGVOLFEDNALAVLNG 120
DB 61 eltyptnaslsflldiovgqvylianhovqvpqlqrlriyrgvolfednyalavldng 120
QY 121 DLNMTPTVPGASPGELRELQRLSLTEILKGVLIQRLPOLCYODTIIMKDFHKNOLA 180
DB 121 dlnmtptvpgaspgelrelqrlslteilkgvliqrlpolcycodtiimkdfhknola 180
QY 121 dlnmtptvpgaspgelrelqrlslteilkgvliqrlpolcycodtiimkdfhknola 180
QY 181 LTLIDTNSRACHPCSPMKGSRGWESSEDCQSLRTVACAGCARCKGPLPTDCHEQC 240
DB 181 ltlidtnsrachpcspmkgsrgwessedcqsllrtvcagcarckgplptdccheqc 240
QY 181 ltlidtnsrachpcspmkgsrgwessedcqsllrtvcagcarckgplptdccheqc 240
QY 241 AAGCTGPKHSDDLACLFHNSGICELHCPALVTYNTDFESMPNEGRITFGASCYTACP 300
DB 241 aagctgpkhsddlacLFHNSGICELHCPALVTYNTDFESMPNEGRITFGASCYTACP 300
QY 241 aagctgpkhsddlacLFHNSGICELHCPALVTYNTDFESMPNEGRITFGASCYTACP 300
QY 301 YVYLSIDVGSCTLVCPPLHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPAAVPVP 355
DB 301 yvylsidvgscTLVCPPLHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPAAVPVP 355

|||||
Db 301 ynylstvsgctlvcp|hngvtaedgtqrcckskpcarvcyglgmeh|revravtsan 360
QY 356 LRMQPG--PAHPVLSFLRSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
Db 361 lgefagckkifgslaf|psfdgdpasnt---ap|qpedq|ygfetleeltgy|lyisawpd 417
QY 406 --PDAHVAVNLSRREG 419
Db 418 slpdlsvfqn|qvtrg 433

RESULT 6
AAW19764 standard; Protein; 782 AA.
ID AAW19764:
AC AAW19764:
XX 17-SEP-1997 (first entry)
DT
XX Her2-GM-CSF immunostimulant fusion protein.
DE
XX Her2-GM-CSF; granulocyte macrophage colony stimulating factor;
KW growth factor receptor; oncogene; immunostimulant; cancer;
KW therapy.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Protein 1..653
FT /label= Her2
FT Peptide 654..655
FT /label= Linker
FT Protein 656..782
FT /label= GM-CSF
XX WO9724438-A1.
PN 10-JUL-1997.
PD 23-DEC-1996; 96WO-US20241.
PF 28-DEC-1995; 95US-0579823.
PR (ACT1-) ACTIVATED CELL THERAPY INC.
XX Laus R, Ruegg CL, Wu H;
PI WPI; 1997-363674/33.
DR N-PSDB; AAT72725.
XX Potent APC that activates T-cells to give multivalent cellular
PT immune response - can also induce a cytotoxic T-cell response in a
PT vertebrate subject
PS Disclosure: Fig 8; 45pp; English.
XX A fusion protein (AAW19764) comprises Her2 (a growth factor receptor
CC that is overexpressed in breast, ovarian can other cancer cells)
CC and granulocyte-macrophage colony stimulating factor (GM-CSF). It
CC is the expression product of a nucleic acid molecule (AAT72725)
CC prep. by PCR amplification of Her2 cDNA from a breast cancer cell
CC line and fusion to GM-CSF cDNA. Fusion expression vectors can be
CC used to transfect mammalian and insect cells. The Her2-GM-CSF
CC fusion protein is used to generate anti-Her2 immunity. Tumour
CC cells are eliminated by cytotoxic T lymphocytes activated in vivo
CC or in vitro by exposure to antigen-presenting cells exposed to the
CC fusion protein.
SQ Sequence 782 AA:

Query Match 82.1%; Score 1878; DB 18; Length 782;

Best Local Similarity 83.0%; Pred. No. 2,2e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRGGLLALLPPRAASTQVCTGDMKRLRPASETHLDMIRHYGOCQVYQGL 60
Db 1 melaalcrngllallppraastqvc|tdmkrlrpaspethldm|rhygqcvyvgnl 60
QY 61 ELTYLPTNASLSFLQDIOGVGVLIANQVROVPLQRLRIYRGTOLEFEDNALAVLNG 120
Db 61 eltylptnaslsflqdigevgyvliahnqvrvp|qlr|lrvygc|fednalavldng 120
QY 121 DPLNMTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTIWKDIFHRNNOLA 180
Db 121 dplnmttpvtgasp|glrel|qlrst|ellk|ggvli|qrnp|lcyqdt|lwkdt|fhrnq|la 180
QY 181 LFLIDITNRSRACHPCSPMKSGRCWGESEDDQSLRTYVACAGCARCKPLPTDCHEQC 240
Db 181 lfliditnrsrachpcspmksgrcwgesedcgs|lrtlvcag|carckp|lptdc|heqc 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
Db 241 aagctgpkhnsdclachfnhsgicelhcpalvtyn|tdf|esmpn|degr|ytf|gascv|ta|cp 300
QY 301 YNYLSTVSGCTLVCP|LHNOEYTAEDGTQRCCKSKPCAR-----GTHSLP|PPAAV|PVP 355
Db 301 ynylstvsgctlvcp|hngvtaedgtqrcckskpcarvcyglgmeh|revravtsan 360
QY 356 LRMQPG--PAHPVLSFLRSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
Db 361 lgefagckkifgslaf|psfdgdpasnt---ap|qpedq|ygfetleeltgy|lyisawpd 417
QY 406 --PDAHVAVNLSRREG 419
Db 418 slpdlsvfqn|qvtrg 433

RESULT 7
AAB21203 standard; protein; 919 AA.
ID AAB21203:
AC AAB21203:
XX 12-JAN-2001 (first entry)
DT
XX Human HER-2/neu fusion protein.
DE
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer; fusion protein.
XX Homo sapiens.
OS Synthetic.
OS WO200044899-A1.
PN 03-AUG-2000.
PD 28-JAN-2000; 2000WO-US02164.
PF 29-JAN-1999; 99US-0117976.
PR (CORI-) CORIXA CORP.
PA (SMK) SMITHKLINE BEECHAM.
XX Cheever MA, Gheysen D;
PI WPI; 2000-505976/45.
DR
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
PS Claim 2; Fig 12; 128pp; English.

XX The present sequence is a fusion protein comprising the extracellular
CC domain and the phosphorylation domain of the human HER-2/neu protein.
CC HER-2/neu is a member of the tyrosine kinase family of receptor-like
CC glycoproteins and shows homology to the epidermal growth factor receptor
CC (EGFR). It probably plays a part in cell growth and/or differentiation.
CC The HER-2/neu gene is an oncogene. HER-2/neu fusion proteins may be used
CC to treat or prevent cancer by eliciting or enhancing an immune response
CC to the HER-2/neu protein. They may be used to treat malignancies such as
CC breast, ovarian, colon, lung and prostate cancers, and may be used as an
CC antigen to vaccinate against these neoplasias.

XX Sequence 919 AA:

Query Match 82.1%; Score 1878; DB 21; Length 919;
Best Local Similarity 83.0%; Pred. No. 2,7e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METAAACRMGILLALPPGAASVOVCTGDMKRLRPASPEFHLDMLRHLVGGQVVOGNTL 60
DB 1 metaaactrglllallppgaastvctgdmkrlrpaspethldmlrhllyggcqvvgntl 60
QY 61 ELTYLPTNALSFLDIOEVGVVLIANQVNPQLQRLRIYRGTOLEFEDNTALAVLDNG 120
DB 61 eltylptnaslflqdigevgvvlianhqvrvpqlrlrlyrgtqlfednyalavldng 120
QY 121 DEPLNNTPTVYGASPGGLRELQRLSEILKGGVLIQRNQOLCYODTILMKDIFHKNNOLA 180
DB 121 dplnntptvtygaspgglrelqrlseilkggvlilqrnqolcyodtllmkdihknnqla 180
QY 181 LFLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVCAGACARCKGRLPTDCCHQC 240
DB 181 lflidtnsrachpcspmksgrcwgessecdqslrtvcagacarkgplptdcchqc 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNBEGRYTFGASCVTACP 300
DB 241 aagctgpkhsdclacfhfnsigicelhcpalvtynntdfesmpnbeqyrtfgascvtacp 300
QY 301 YNYLSTDVGSCITLVCPILHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPRAVPVP 355
DB 301 ynylstdvgsctlvcpilhnoevtaedgtqrcekskpcarvcyglgmehllrevratlsan 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFVSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 356 lrmqpg--pahpvlslflrpsmdlvsafovslplaplsptsvpi-----spvsvgrgpd 405
QY 406 igeagckkifgslafipstfdgdpasnt---aplpqeqlyfettleetylylsawpd 417
DB 406 igeagckkifgslafipstfdgdpasnt---aplpqeqlyfettleetylylsawpd 417
QY 418 sldpdlsvfnlqvtrg 433
DB 418 sldpdlsvfnlqvtrg 433

RESULT 8

AAB21208 AAB21208 standard; Protein; 1200 AA.

AC AAB21208;

DT 12-JAN-2001 (first entry)

XX Human HER-2/neu protein.

KM Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; Vaccine;

KM breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX Homo sapiens.

XX MO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIAX CORP.

XX (SMK) SMITHLINE BEECHAM.

XX Cheever MA, Gheysen D;

XX WPI: 2000-505976/45.

XX N-PSDB; AAA89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins

XX useful for vaccinating against breast, ovarian, colon, lung and

XX prostate cancers -

XX Disclosure: Fig 15; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member
CC of the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.

XX Sequence 1200 AA:

Query Match 82.1%; Score 1878; DB 21; Length 1200;
Best Local Similarity 83.0%; Pred. No. 3.8e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METAAACRMGILLALPPGAASVOVCTGDMKRLRPASPEFHLDMLRHLVGGQVVOGNTL 60
DB 1 metaaactrglllallppgaastvctgdmkrlrpaspethldmlrhllyggcqvvgntl 60
QY 61 ELTYLPTNALSFLDIOEVGVVLIANQVNPQLQRLRIYRGTOLEFEDNTALAVLDNG 120
DB 61 eltylptnaslflqdigevgvvlianhqvrvpqlrlrlyrgtqlfednyalavldng 120
QY 121 DEPLNNTPTVYGASPGGLRELQRLSEILKGGVLIQRNQOLCYODTILMKDIFHKNNOLA 180
DB 121 dplnntptvtygaspgglrelqrlseilkggvlilqrnqolcyodtllmkdihknnqla 180
QY 181 LFLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVCAGACARCKGRLPTDCCHQC 240
DB 181 lflidtnsrachpcspmksgrcwgessecdqslrtvcagacarkgplptdcchqc 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNBEGRYTFGASCVTACP 300
DB 241 aagctgpkhsdclacfhfnsigicelhcpalvtynntdfesmpnbeqyrtfgascvtacp 300
QY 301 YNYLSTDVGSCITLVCPILHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPRAVPVP 355
DB 301 ynylstdvgsctlvcpilhnoevtaedgtqrcekskpcarvcyglgmehllrevratlsan 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFVSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 356 lrmqpg--pahpvlslflrpsmdlvsafovslplaplsptsvpi-----spvsvgrgpd 405
QY 406 igeagckkifgslafipstfdgdpasnt---aplpqeqlyfettleetylylsawpd 417
DB 406 igeagckkifgslafipstfdgdpasnt---aplpqeqlyfettleetylylsawpd 417
QY 418 sldpdlsvfnlqvtrg 433
DB 418 sldpdlsvfnlqvtrg 433

RESULT 9

AAW01111 AAW01111 standard; Protein; 1255 AA.

XX

```

AC  AAM01111;
XX
DT  01-JAN-1997 (first entry)
XX
DE  HER-2/neu protein.
XX
KW  HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
RW  breast cancer; ovary cancer; colon cancer; lung cancer;
KM  prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS  Homo sapiens.
XX
FH  Key
FT  Domain
FT  Location/Qualifiers
   /label= Intracellular_domain
   /note= "claimed domain, useful for immunisation"
XX
PN  W09630514-A1.
XX
PD  03-OCT-1996.
XX
PE  28-MAR-1996; 96MO-US01689.
XX
PR  31-MAR-1995; 95US-0414417.
XX
PA  (UNIM ) UNIV WASHINGTON.
XX
PI  Cheever MA, Disls ML;
XX
DR  WPI: 1996-455361/45.
DR  N-PSDB: AAT40739.
XX
PT  DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT  treatment of malignancies with which the HER-2/neu oncogene is
PT  associated
XX
PS  Claim 2; Page 56-61; 71pp; English.
XX
CC  Human HER-2/neu protein (AAM01111), also called p185 or c-erbB2, is
CC  the product of the HER-2/neu oncogene (see also AAT40739). The
CC  protein is over-expressed in various cancers, including breast,
CC  ovarian, colon, lung and prostate. The intracellular domain of the
CC  protein can be used to immunise an animal against a malignancy with
CC  which the oncogene is associated. The polypeptide can be produced
CC  in transformed host cells for use in immunisation. Alternatively,
CC  animal cells are transfected in vivo or ex vivo with a viral vector
CC  that directs expression of the polypeptide.
XX
SQ  Sequence 1255 AA;

Query Match      82.1%; Score 1878; DB 17; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY  1  MELALALRMGLLLALPLPGAASSTOVCSTGDKRLRLPASPEHLDMLRLHYOGCOVQGNL 60
DB  1  melalalrmgl11al1ppgaastvcyctdmk11lpaspehldmlrhlhyogcqvqgnl 60
OY  61  ELTYLPTNASLFLQDIDQEVGYVLIANQVROYPLQRLRLVGTQTFEDNVAALAVLNG 120
DB  61  eltylptnaslflqdidqevgyvlianhgyrpylqrlrlvgtqflednvaalavlng 120
OY  121  DPLNNTPTVTSASRGGLRELQRLSTELKGGVLIQRNPOLCYQDTILMKDIFRKNOLA 180
DB  121  dplnntptvtgaspqglrelqrlsteltllkgvliqrnpqlcyqdtllmkdlfrknqla 180
OY  181  LTLIDTNSRACHPCSPCKSGRCSESEDCOSTLRTVCGAGCARCKGPLPTDCHFCOC 240
DB  181  ltlidtnsrachpcspcksgrcsgesedcgsllrtvcgagcarckgplptdchfcgc 240
OY  241  AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRVYTGASCVTACP 300
DB  241  aagctgpkhsdclacLFHNSGICELHCPALVTYNTDTFESMPNDEGRVYTGASCVTACP 300

```

```

DB  241  aagctgpkhsdclacLFHNSGICELHCPALVTYNTDTFESMPNDEGRVYTGASCVTACP 300
OY  301  YNYLSTDVSGCTIVCPLEHNOEVTAEEDGROCKECSKPCAR-----GTHSLPRAPAVVP 355
DB  301  ynylstdvsgctivcplnhgevtadgtrceksckpcarvcyglqnehlfrevavtsan 360
OY  356  LRMPG--PAHPVLSFLRPSMDLVSFAFYSLPLAPLSPISVP1-----SPVSGRGCPD 405
DB  361  lqefagckkifgsiaflpfesfdgdpasnt---aplpqeqlyvfeleltgylylsawpd 417
OY  406  --PDAAHVAVNLSRYEG 419
DB  418  slpdlstvfgnlqvltqg 433

RESULT 10
AAM92406
ID  AAM92406 standard; Protein: 1255 AA.
XX
AC  AAM92406;
XX
DT  21-APR-1999 (first entry)
XX
DE  Human HER-2/neu oncogene protein.
XX
KW  HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KM  malignancy; treatment; tumour.
XX
OS  Homo sapiens.
XX
FH  Key
FT  Region
FT  Location/Qualifiers
   /note= "region which elicits immune response"
XX
PN  US5869445-A.
XX
PD  09-FEB-1999.
XX
PE  01-APR-1996; 96US-0625101.
XX
PR  01-APR-1996; 96US-0625101.
PR  17-MAR-1993; 93US-0033644.
PR  12-AUG-1993; 93US-0106112.
PR  31-MAR-1995; 95US-0414417.
XX
PA  (UNIM ) UNIV WASHINGTON.
XX
PI  Cheever MA, Disls ML;
XX
DR  WPI: 1999-152835/13.
DR  N-PSDB: AAX01912.
XX
PT  Use of HER-2/neu polypeptides - for eliciting an immune response to
PT  an HER-2/neu associated malignancy, particularly for treating or
PT  preventing tumours
XX
PS  Claim 3; Column 31-38; 26pp; English.
XX
CC  This sequence represents the human HER-2/neu oncogene protein. A fragment
CC  of this protein is used in a method for eliciting or enhancing an immune
CC  response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC  B cells to produce an immune response to the HER-2/neu protein. The
CC  method can be used for immunisation against a malignancy in which the
CC  HER-2/neu oncogene is associated and in the treatment of an existing
CC  tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ  Sequence 1255 AA;

Query Match      82.1%; Score 1878; DB 20; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```

```

Oy 1 MELALCRWGLLLALLPPGAASCTGCTGDMKRLRPASPEHLDMLRHLXGCGVVGNL 60
    |||
Db 1 melalcrwgl11allppgaastgctgtdmkrlrpspehldmlrhllygqcyvggnl 60
Oy 61 ELTYLPTNASSFLDODIEVOGYVLIANQVROVPLQRLRIYRGTOLEFDNYALAVDNG 120
    |||
Db 61 eltylptnassflqddievggyvlianhqvrvplqlrlirvrgtqlfednyalavldng 120
Oy 121 DPLNNTPTVTGASPGGLRELQRLSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
    |||
Db 121 dplnntcptvgaspgglrelqlrlsltelkgyvliqrnpqlcygdtllwkdlfhknq1a 180
Oy 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRTVACAGCARCGPLPTDCHEQC 240
    |||
Db 181 ltlidtnsrachpcspcksgrcwgessecdqsltrtvacagcarckgplptdccheqc 240
Oy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDFESMPNDEGRYTGASCVTACP 300
    |||
Db 241 aagctgpkhsdclaclfhnhsgicelhcpalvtlyntdfesmpnpegrytlfgascvtlacp 300
Oy 301 YNYLSTDVSGCTLVCPLEHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAAYVP 355
    |||
Db 301 ynylstdvsgctlvcpplhnoevtaedgtqrcekskpcarvcyglgmehlrevravtsan 360
Oy 356 LRMOPG--PAHPVLSFLRPSMDVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
    |||
Db 361 lqefagckkifgslafllpesfdgdpasnt---apllqpeqlqvfeleeltyglyisawpd 417
Oy 406 --PDAHVAVNLSRYEG 419
    |||
Db 418 slpdlsvfqnqlqvtrg 433

```

RESULT 11

```

AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human: HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
PN WO200044899-A1.
XX
PD 03-AUG-2000.
XX
PE 28-JAN-2000; 2000WO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX
PA (CORI-) CORIAX CORP
PA (SMK ) SMITHKLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX
DR WPI: 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows

```

```

CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;

```

```

Query Match      82.1%  Score 1878; DB:21; Length 1255;
Best Local Similarity 83.0%  Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```

```

Oy 1 MELALCRWGLLLALLPPGAASCTGCTGDMKRLRPASPEHLDMLRHLXGCGVVGNL 60
    |||
Db 1 melalcrwgl11allppgaastgctgtdmkrlrpspehldmlrhllygqcyvggnl 60
Oy 61 ELTYLPTNASSFLDODIEVOGYVLIANQVROVPLQRLRIYRGTOLEFDNYALAVDNG 120
    |||
Db 61 eltylptnassflqddievggyvlianhqvrvplqlrlirvrgtqlfednyalavldng 120
Oy 121 DPLNNTPTVTGASPGGLRELQRLSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
    |||
Db 121 dplnntcptvgaspgglrelqlrlsltelkgyvliqrnpqlcygdtllwkdlfhknq1a 180
Oy 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRTVACAGCARCGPLPTDCHEQC 240
    |||
Db 181 ltlidtnsrachpcspcksgrcwgessecdqsltrtvacagcarckgplptdccheqc 240
Oy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDFESMPNDEGRYTGASCVTACP 300
    |||
Db 241 aagctgpkhsdclaclfhnhsgicelhcpalvtlyntdfesmpnpegrytlfgascvtlacp 300
Oy 301 YNYLSTDVSGCTLVCPLEHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAAYVP 355
    |||
Db 301 ynylstdvsgctlvcpplhnoevtaedgtqrcekskpcarvcyglgmehlrevravtsan 360
Oy 356 LRMOPG--PAHPVLSFLRPSMDVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
    |||
Db 361 lqefagckkifgslafllpesfdgdpasnt---apllqpeqlqvfeleeltyglyisawpd 417
Oy 406 --PDAHVAVNLSRYEG 419
    |||
Db 418 slpdlsvfqnqlqvtrg 433

```

RESULT 12

```

AA84780
ID AA84780 standard; protein; 1255 AA.
XX
AC AA84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPICE erbB-2 receptor protein.
XX
KW SPICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
OS Homo sapiens.
XX
PN WO200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX

```

PR 02-OCT-1998; 98US-0165192.
XX (UYMC-) UNIV MCMMASTER.
XX
PI Muller WJ, Siegel PM;
XX WPI: 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
XX erbB-2, inhibitors of the protein are useful for treatment of cancer
PS
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match 82.1%; Score 1878; DB 21; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MEIAALCWMGLLALLPGCASTOVCTGDMKRLPASPETHLDMRLHYGCGVVGNTL 60
DB 1 melaalcwmgllallppgaastqvcgtmdkrlpaspethldmrlhygcvvgvgnl 60
QY 61 ERTYPTNASTSLFDIOIQVQGVLIANQVROVPLQRLRYRGTOLEFDNTALAVLDNG 120
DB 61 ertyptnastslfdioiqvqgvlianhqvrovplqrlryrgtolfednyalavldng 120
QY 121 DPLNNTPTVGASPGGLRLOLRSLTEILKGVLIORNPOLCYODITLKKDIFHKNNOLA 180
DB 121 dplnntptvgaspgglrlolrslteilkgvliornpolicyoditlkkdifhknola 180
QY 181 LTLIDTNRSRACHPCSPMKGSRGSESEDCOSLTRVTCAGACARCKGPLPTDCHEQC 240
DB 181 ltlidtnrsrachpcspmkgsrgsesedcosltrvtcagacarckgplptdccheqc 240
QY 241 AAGCGPKKSDCLACLFHNSGICELHCPALVYNTDFESMPNREGRTFAGASCVTACP 300
DB 241 aagcgpkksdclacLFHNSGICELHCPALVYNTDFESMPNREGRTFAGASCVTACP 300
QY 301 YVYLSDVGSCTLVCPRIHQEYTAEDGTORCEKSKPCAR-----GTHSLRPAAPVVP 355
DB 301 yvylsdvgsctlvcprihqeytaedgtorcekscpcar-----gthslrpaavvp 355
QY 356 LRMQRC--PAHPVLSFLRPSMDLVSAFVSLPLAPLSPTSVPI-----SPVSVGRPD 405
DB 356 lrmqrc--pahpvlSFLRPSMDLVSAFVSLPLAPLSPTSVPI-----spvsvgrpd 405
QY 406 --PDAHVAVNLSRYEG 419
DB 406 --pdahvaVNLSRYEG 419
QY 418 slpdlsvfqnqlvyrq 433
DB 418 slpdlsvfqnqlvyrq 433

RESULT 13
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;

XX 10-AUG-2000 (first entry)
DT
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
XX Key
FH Domain
FT
FT
FT Location/Qualifiers
FT 1..173
FT /label= "N-terminal
FT /note= "mature polypeptide"
FT 5..25
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 59..73
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 103..117
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 149..163
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 174..323
FT /label= Cysteine_rich_domain
FT 210..224
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 250..264
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 324..483
FT /label= Ligand_binding_domain
FT 325..339
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 369..383
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 465..479
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 484..623
FT /label= Cysteine_rich_domain
FT 579..593
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 624..654
FT /label= Transmembrane_domain
FT 632..652
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 653..667
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 655..1010
FT /label= Tyrosine_kinase_domain
FT 661..675
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 695..709
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 710..730
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 1011..1235
FT /label= C-terminal_domain
XX Domain

PN WO20020027-A2.
XX
PD 13-APR-2000.
XX
PF 05-OCT-1999; 99WO-DK00525.
XX
PR 05-OCT-1998; 98BK-0001261.
PR 20-OCT-1998; 98US-0105011.
XX
PA (MEBI-) M & E BIOTECH AS.
XX
PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I,
PI Gautam A, Birk P, Karlsson G;
XX
DR WPI: 2000-349917/30.
XX N-PSDB; AAA09455.
XX
XX Inducing immune responses to weakly immunogenic, tumor associated
PT peptide antigens for the treatment of breast and prostate cancer
XX
PS Claim 62; Page 193-198; 220pp; English.
XX
XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC response. Subdominant CTL epitopes, antibody binding regions and
CC cysteine residues involved in disulfide bonds are preserved in the
CC immunogenized forms. Regions suitable for the insertion of foreign T
CC helper epitopes were identified (see features table). The method
CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 1255 AA;
Query Match 82.1%; Score 1878; DB 21; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
OY 1 MELAALCRWGLLALLPRAASTQVCTGDMKRLPASPETHLDMRLHLYOGCQVVGNTL 60
DB 1 melaaalcrwglallprraastqvtctgtdmkrlrpsapethldmrlhlygqcvvgvgnl 60
OY 61 ELTYLPFNASLSFLQDIOGVGVYLLAHNOVROPLORLRIYRGTOLEFDNVALAVLNDG 120
DB 61 eltylpfnaslsflqdigvgvyllahngvrvpqlrlyrgtqlfednyalavldng 120
OY 121 DPLNNTPTVTGASPGRLRELQLRSLTEILKGVLIQIRNPQLCYQDILFWDLHKNQOLA 180
DB 121 dplnntptvtgaspgrlrelqlrslteilkgvliqirnpqlcyqdcilfwdlhknqola 180
OY 181 LFLIDNRRACHPCSPMKGRSGWESSDQSLRFTVCAGCARGKGLPDCHECC 240
DB 181 lflidnrrachpcspmkgrsgwessedqslrftvcagcargkglpdcchecc 240
OY 241 AAGCTGPKRHSDDLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCTYACP 300
DB 241 aagctgpkrsdclachfnhsgicelhcpalvynntdfesmpnpgrytfgasctyacr 300
OY 301 YNTLSTDVSCSTLVCPLAHNOEVTAEDEGTORCEKSPCAR----GTHSLPRAAPVYP 355
DB 301 yntlstdvscstlvcplahnoevtaedegtorcekspcar-----gthslprraavyp 355
DB 301 ynlstdvscstlvcplahngvtaedgtqrcckspcarvcyglgmehlravrtsan 360

OY 356 LRMQPG--PAHPVLSFLRSPMDLSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 361 lqefagckkifsglsflrpsfdgdpasnt---aplqpeqlgvfclleatlylyisawpd 417
OY 406 --PDAHVAVNLSRYEG 419
DB 418 slpdlsvfnglqvtrg 433
RESULT 14
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
XX AAB85458;
AC
XX 25-SEP-2001 (first entry)
DT
XX
DE Human HER-2/neu protein.
XX
KM Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
XX Homo sapiens.
XX
XX WO200153463-A2.
PD 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US01850.
PF
XX 21-JAN-2000; 2000US-0177545.
PR
XX (CORI-) CORIXA CORP.
PA
XX Cheever MA, Hand-Zimmermann S;
PI
XX WPI: 2001-476112/51.
DR N-PSDB; AAH23392.
XX
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer
XX
XX
PS Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
Query Match 82.1%; Score 1878; DB 22; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
OY 1 MELAALCRWGLLALLPRAASTQVCTGDMKRLPASPETHLDMRLHLYOGCQVVGNTL 60
DB 1 melaaalcrwglallprraastqvtctgtdmkrlrpsapethldmrlhlygqcvvgvgnl 60
OY 61 ELTYLPFNASLSFLQDIOGVGVYLLAHNOVROPLORLRIYRGTOLEFDNVALAVLNDG 120
DB 61 eltylpfnaslsflqdigvgvyllahngvrvpqlrlyrgtqlfednyalavldng 120
OY 121 DPLNNTPTVTGASPGRLRELQLRSLTEILKGVLIQIRNPQLCYQDILFWDLHKNQOLA 180
DB 121 dplnntptvtgaspgrlrelqlrslteilkgvliqirnpqlcyqdcilfwdlhknqola 180

Db 121 dplnltptvgaspqglrelqlrsltelkqgvlqgrnplcygdltlwkdlfhknqqla 180
QY 181 LFLITNRSRACHPCSPMKGRGCGESSEDCQSLTRTYCAGGCAKRCCKPLPTDCHEQC 240
Db 181 llllcltnsrachpcspmkgrcgessedcqslltrvcaggcarckypjpldccheqc 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRRTFGASCVTACP 300
Db 241 aagctgpkhsdclaclhfnhsigicelhcpcalvtlyntdfesmpnpegryltfgascvtacp 300
QY 301 YNYLSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSPCAR-----GTHSLPRPAVPVP 355
Db 301 ynylstdvsgctlvcpplhnoevtaedgtqrcckspcarvegylgmehlrevravtsan 360
QY 356 LRMQPG--PAHNVLSFLRSMDLVSAFYSLLPAPLSPTSVPT-----SPVSYGRGPD 405
Db 361 lqefagckkllfgslafllpeditgdpasnt---apllqpeqlqvfeclteeltgylisawpd 417
QY 406 --PDAHVAVNLSRYEG 419
Db 418 slpdlsvfqnlgvtrg 433

RESULT 15
AAG88267
ID AAG88267 standard; Protein: 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW Immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Cells E;
PI Keogh E;
XX
DR WPI: 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
PS Disclosure: Page 15; 1999p: English.

CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and

CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected pathological side effects caused by
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.

SO Sequence 1255 AA;

Query Match 82.1%; Score 1878; DB 22; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELALCRKGLLALLPFGAATQVCTGTDMLRPAFETHLDMRLHYGCGYVQCNL 60
Db 1 melalcrwgllallppgaastqvcyctldmklrllpaspetldmlrhllygqcyvqgnl 60
QY 61 ELTYLPTNASLSFLDIOEVQGVLIHNOVQVPLQRLRIYRGTOLPEDNYALAVLNG 120
Db 61 eltylptnaslsfllqdvqvgvyliahnqvtrpqlrltrivrgqllfednyalavldng 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDIFHRNNOLA 180
Db 121 dplnltptvgaspqglrelqlrsltelkqgvlqgrnplcygdltlwkdlfhknqqla 180
QY 181 LFLITNRSRACHPCSPMKGRGCGESSEDCQSLTRTYCAGGCAKRCCKPLPTDCHEQC 240
Db 181 llllcltnsrachpcspmkgrcgessedcqslltrvcaggcarckypjpldccheqc 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRRTFGASCVTACP 300
Db 241 aagctgpkhsdclaclhfnhsigicelhcpcalvtlyntdfesmpnpegryltfgascvtacp 300
QY 301 YNYLSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSPCAR-----GTHSLPRPAVPVP 355
Db 301 ynylstdvsgctlvcpplhnoevtaedgtqrcckspcarvegylgmehlrevravtsan 360
QY 356 LRMQPG--PAHNVLSFLRSMDLVSAFYSLLPAPLSPTSVPT-----SPVSYGRGPD 405
Db 361 lqefagckkllfgslafllpeditgdpasnt---apllqpeqlqvfeclteeltgylisawpd 417
QY 406 --PDAHVAVNLSRYEG 419
Db 418 slpdlsvfqnlgvtrg 433

Search completed: April 11, 2002, 09:29:11
Job time: 53 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:30:07 ; Search time 25.01 Seconds
(without alignments)
1276.175 Million cell updates/sec

Title: US-09-234-208B-2
Perfect score: 2287
Sequence: 1 METALCRMGLLALPPGA.....VGRGPPDAHVAVNLSRREG 419

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	82.1	1255	1 A24571	protein-tyrosine k
2	1608.5	70.3	1260	1 TVRTNU	protein-tyrosine k
3	1571	68.7	1254	2 148161	p-185 precursor -
4	796.5	34.8	1223	1 TVCHLV	epidermal growth f
5	794	34.7	527	2 A42032	epidermal growth f
6	793	34.7	1210	1 GQHUE	epidermal growth f
7	789	34.5	1210	2 A53183	epidermal growth f
8	781.5	34.2	644	2 A36325	epidermal growth f
9	775	32.9	1308	2 A47253	epidermal growth f
10	735.5	32.2	1342	2 A36223	epidermal growth f
11	692	30.3	1339	2 JC4387	kinase-related tra
12	681.5	29.8	1166	1 S06142	epidermal growth k
13	575.5	25.2	843	2 A27131	epidermal growth f
14	419	18.3	1374	2 S70712	epidermal growth f
15	416	18.2	1369	2 S70713	protein-tyrosine k
16	415	18.1	1330	1 GQFE	epidermal growth f
17	363.5	15.9	366	2 D45558	epidermal growth f
18	363.5	15.9	1717	1 A45558	epidermal growth f
19	342.5	15.0	1363	2 T43220	insulin-like growt
20	331	14.5	333	2 B45558	epidermal growth f
21	331	14.5	342	2 C45558	epidermal growth f
22	294	12.9	2101	2 S57245	insulin receptor f
23	294	12.9	2148	2 A56081	insulin receptor f
24	291	12.7	1477	2 T18534	protein-tyrosine k
25	278	12.2	1300	2 A36502	insulin receptor-r
26	272	11.9	540	2 B47417	insulin receptor-r
27	269.5	11.8	1607	2 T43212	insulin-like growt
28	263.5	11.5	1382	1 INHUR	insulin receptor p
29	263.5	11.5	1383	2 A36080	insulin receptor p

30	261	11.4	1372	2 A34157	insulin receptor p
31	258	11.3	1390	2 T30346	insulin receptor -
32	257.5	11.3	1367	1 IGHUR1	insulin-like growt
33	252.5	11.0	1371	2 A33837	insulin-like growt
34	249	10.9	1268	2 B36502	insulin receptor-r
35	235	10.3	329	2 A48805	tyrosine kinase re
36	226.5	9.9	183	2 JH0803	insulin receptor h
37	203	8.9	1846	2 T42047	insulin receptor h
38	160.5	7.0	1299	2 T43251	furin (EC 3.4.21.7
39	144	6.3	1548	2 S34583	serine proteinase
40	142	6.2	1959	1 AGRT	agrin - rat
41	136.5	6.0	1111	2 T26972	hypothetical prote
42	133.5	5.8	915	1 A48225	subtilisin-like pr
43	132	5.8	1574	2 T13954	MEGF6 protein - ra
44	131.5	5.7	915	2 B48225	probable proteol
45	131.5	5.7	915	2 JC6148	subtilisin-like pr

ALIGNMENTS

RESULT 1
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein e
C:Species: Homo sapiens (man)
C>Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T
Nature 319, 230-234, 1986
A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A:Reference number: A24571; MUID:86118663
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-references: GB:X03363; NID:931197; PIDN:CAA27060.1; PID:931198
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid
A:Reference number: A25491; MUID:86016729
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-references: GB:M1767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro
A:Reference number: A44188; MUID:86070181
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517; 'RALI', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-references: GB:M11730; NID:g183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597
A:Accession: I59509
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 832-909 <REX>
A:Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio
A:Reference number: I57622; MUID:87286898
A:Accession: I57622
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AAA58637.1; PID:9553332
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C:Genetics: GDB:ERBB2; NGU; HER-2
A:Gene: GDB:ERBB2; NGU; HER-2
A:Cross-references: GDB:120613; OMIM:164870
A:Map position: 17q21.1-17q21.1
A:Introns: 25/1; 75/3; 147/1; 883/3
A>Note: the list of introns is incomplete
C:Function: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
A:Description: epidermal growth factor receptor; protein kinase homology
C:Superfamily: epidermal growth factor receptor; glycoprotein; phosphoprotein; phosphatase
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-125/Domain: protein-tyrosine kinase erbB #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68/124,187,253,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 82.1%; Score 1878; DB 1; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4,2e-121;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY 1 MELAALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHLYOGCQVQGNL 60
DB 1 MELAALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHLYOGCQVQGNL 60
OY 61 ELTYLPNTASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 120
DB 61 ELTYLPNTASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 120
OY 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQNRPOLCYQDITLMKDIFFKNNQLA 180
DB 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQNRPOLCYQDITLMKDIFFKNNQLA 180
OY 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQNRPOLCYQDITLMKDIFFKNNQLA 180
DB 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQNRPOLCYQDITLMKDIFFKNNQLA 180
OY 181 LTLIDITNRSRACHPCSPCKGRCWGESSEDCQSLRTVCAGCARKGRLPTDCCHQ 240
DB 181 LTLIDITNRSRACHPCSPCKGRCWGESSEDCQSLRTVCAGCARKGRLPTDCCHQ 240
OY 181 LTLIDITNRSRACHPCSPCKGRCWGESSEDCQSLRTVCAGCARKGRLPTDCCHQ 240
DB 181 LTLIDITNRSRACHPCSPCKGRCWGESSEDCQSLRTVCAGCARKGRLPTDCCHQ 240
OY 241 AAGCTGPRKHSDDLCLHFNHSGICELHCPALVYNTDFESMPNDEGRYTFGASCYTAC 300
DB 241 AAGCTGPRKHSDDLCLHFNHSGICELHCPALVYNTDFESMPNDEGRYTFGASCYTAC 300
OY 301 YNLTSLDVGSCTIVCPNLHNOEVAEDGTQRCCKSPCAR-----GTHSLPRAVP 355
DB 301 YNLTSLDVGSCTIVCPNLHNOEVAEDGTQRCCKSPCAR-----GTHSLPRAVP 355
OY 301 YNLTSLDVGSCTIVCPNLHNOEVAEDGTQRCCKSPCAR-----GTHSLPRAVP 355
DB 301 YNLTSLDVGSCTIVCPNLHNOEVAEDGTQRCCKSPCAR-----GTHSLPRAVP 355
OY 356 LRMQPC--PAHPLVSLFRPSMDVSAFSLPLASPTSVPI-----SPVSVGRGD 405
DB 356 LRMQPC--PAHPLVSLFRPSMDVSAFSLPLASPTSVPI-----SPVSVGRGD 405
OY 361 IQEFACCKITFGSLAFPLPESFDGDPASNT---APLOPEQLQVETLEITGYLISAMP 417
DB 361 IQEFACCKITFGSLAFPLPESFDGDPASNT---APLOPEQLQVETLEITGYLISAMP 417
OY 406 --PDAAHVAANLSRYREG 419
DB 418 SLPDLSPFQNLQYIRG 433

RESULT 2
TYRTNU
protein-tyrosine kinase (BC 2.7.1.112) neu precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Barbmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986

A>Title: The neu oncogene encodes an epidermal growth factor receptor-related protein
A:Reference number: A24562; MUID:86118662
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Colhe
Carcinogenesis 12, 1975-1978, 1991
A>Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals a
2-thiazolylformamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293
A:Accession: A61204
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 637-663, 'V', 665-702 <MAS>
A>Note: authors translated the codon GCA for residue 25 as Val
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Domain: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TM>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 70.3%; Score 1608.5; DB 1; Length 1260;
Best Local Similarity 85.0%; Pred. No. 1.3e-102;
Matches 294; Conservative 15; Mismatches 36; Indels 1; Gaps 1;

OY 1 MELAALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHLYOGCQVQGNL 60
DB 4 MELAALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHLYOGCQVQGNL 63
OY 61 ELTYLPNTASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 120
DB 61 ELTYLPNTASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 120
OY 64 ELTYVPANASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 123
DB 64 ELTYVPANASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 123
OY 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQNRPOLCYQDITLMKDIFFKNNQ 179
DB 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQNRPOLCYQDITLMKDIFFKNNQ 179
OY 124 DQDNVNAASTPRTFEGLEQLRSLTEILKGVLIQNRPOLCYQDITLMKDIFFKNNQ 183
DB 124 DQDNVNAASTPRTFEGLEQLRSLTEILKGVLIQNRPOLCYQDITLMKDIFFKNNQ 183
OY 180 ALTLIDITNRSRACHPCSPCKGRCWGESSEDCQSLRTVCAGCARKGRLPTDCCHQ 239
DB 180 ALTLIDITNRSRACHPCSPCKGRCWGESSEDCQSLRTVCAGCARKGRLPTDCCHQ 239
OY 184 AAVDIDITNRSRACHPCAPKCKDNHGWGESPEDCQLITGTCGCAKCGRLPTDCCHQ 243
DB 184 AAVDIDITNRSRACHPCAPKCKDNHGWGESPEDCQLITGTCGCAKCGRLPTDCCHQ 243
OY 240 CAAGCTGPRKHSDDLCLHFNHSGICELHCPALVYNTDFESMPNDEGRYTFGASCYTAC 299
DB 240 CAAGCTGPRKHSDDLCLHFNHSGICELHCPALVYNTDFESMPNDEGRYTFGASCYTAC 299
OY 244 CAAGCTGPRKHSDDLCLHFNHSGICELHCPALVYNTDFESMPNDEGRYTFGASCYTAC 303
DB 244 CAAGCTGPRKHSDDLCLHFNHSGICELHCPALVYNTDFESMPNDEGRYTFGASCYTAC 303
OY 300 PNYLTSLDVGSCTIVCPNLHNOEVAEDGTQRCCKSPCARSTHSL 345
DB 304 PNYLTSLDVGSCTIVCPNLHNOEVAEDGTQRCCKSPCARSTHSL 349

RESULT 3
148161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: 148161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishika
Gene 140, 251-255, 1994
A>Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: 148161; MUID:94193007
A:Accession: 148161
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:g493236; PIDN:BA03801.1; PID:g747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 68.7%; Score 1571; DB 2; Length 1254;
Best Local Similarity 74.4%; Pred. No. 4.6e-100;
Matches 299; Conservative 26; Mismatches 67; Indels 10; Gaps 3;

QY 1 METALCRNGLLALLPPEASTQVCTGDMKRLPASPEHLDMLRHLXGCGVQGNL 60
DB 1 MELAAACGGGLLALLSPASGCTGDMKRLPASPEHLDIVRHLXGCGVQGNL 60
QY 61 ELTYLPTNASTLFLDIOEVGVYVLIANQVROVPLQRLRIYRGTOLEFEDNVALAVLDNG 120
DB 61 ELTYLRANATLSFLDIOEVGYMLIANSQVRAHVPLORLRIYRGTOLEFEDKITALAVLDNR 120
QY 121 DELNMTPTVVGASPGGLRELQRLSLTEILKGGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
DB 121 DELDNVTTATGRTPEGLRELQRLSLTEILKGGVLIIRGNPOLCYODTILMKDVERKNNOLA 180
QY 181 LRLIDNRARCPSPCKSGRCWGESEDDQSLRTYVACAGCARGKPLTDCHEQC 240
DB 181 PVDIDNRARCPCAPACKDNHCWASPEDCQTLGITAPRAVPAARARLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELCPALVTYNTDFESMPNPEGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELCPALVTYNTDFESMPNPEGRYTFGASCVTAC 300
QY 301 VNYLSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSPCAR----GTHSLRPPAAVVP 355
DB 301 VNYLSTGVSGCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCGMELHNGARAITSAN 360
QY 356 LHMOPG--PAHVLSTLRSPMDLVSAFYSLPLAPISPTSVPI 395
DB 361 IDPAGCKKIFGSLAPLPESFD--GNPSSGIAPLPLEQLQV 399

RESULT 4
TWOCHV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Latx, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A:Reference number: A27720; MUID:88261272
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilisen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NII>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>

F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: Transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136-202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent)
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #s

Query Match 34.8%; Score 796.5; DB 1; Length 1223;
Best Local Similarity 45.5%; Pred. No. 5.6e-47;
Matches 155; Conservative 55; Mismatches 112; Indels 19; Gaps 7;

QY 8 RMGLLALLPPEA-----STQVCTGDMKRLPASPEHLDMLRHLXGCGVQGNL 61
DB 13 KGAVALVLLLCVALLCSAVEKKVCGGINNKLTQGLGHEDHFTSLQRMNCEVLSNLE 72
QY 62 LTYLPTNASTLFLDIOEVGVYVLIANQVROVPLQRLRIYRGTOLEFEDNVALAVLDNGD 121
DB 73 IYVEHNRLDPLFKTIOEVAGVYVLIANQVDTPLENLIQINGVLYDONSFAVALSYNH 132
QY 122 PLNMTPTVVGASPGGLRELQRLSLTEILKGGVLIQIRNPOLCYODTILMKDIFHKNNOLA 181
DB 133 -MNKTQ-----GLRELPMKRLEIINGVAKISNNPRLCMDDVNLNDIIDSRSK-PL 182
QY 182 TLID-TNRSRACHPSPCKSGRCWGESEDDQSLRTYVACAGCA-RCKGPLPTDCHEQ 239
DB 183 TVLDRAINSLSCKPKHPNCTEDHCAGBQNCQTLTKYCAQCCGRCGRKPKSPCCHQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSGICELCPALVTYNTDFESMPNPEGRYTFGASCVTAC 299
DB 243 CAAGCTGPRESDCLACRFKRDATCKDPCPLVLYNQMDVNPBGRKYSRGATVREC 302
QY 300 PNYLSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSPCAR 340
DB 303 PHNYVYTDHSGCVRSCNPTDYEY-ENGVRCKCKDGLOSK 342

RESULT 5
A42032
epidermal growth factor receptor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 18-Jun-1999
C:Accession: A42032
R:Flickinger, T.W.; Mahle, N.J.; Kung, H.J.
Mol. Cell. Biol. 12, 883-893, 1992
A:Title: An alternatively processed mRNA from the avian c-erbB gene encodes a soluble
A:Reference number: A42032; MUID:92123214
A:Accession: A42032
A:Molecule type: mRNA
A:Status: preliminary
A:Residues: 1-527 <FLI>
A:Cross-references: GB:M77637; NID:g211737; PIDN:AAA48759.1; PID:g211738
A:Experimental source: Liver
A:Note: sequence extracted from NCBI backbone (NCBI:76892, NCBI:76893)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor

Query Match 34.7%; Score 794; DB 2; Length 527;
Best Local Similarity 46.0%; Pred. No. 3.4e-47;
Matches 155; Conservative 55; Mismatches 109; Indels 18; Gaps 7;

QY 11 LIALLLPPEAST-----QVCTGDMKRLPASPEHLDMLRHLXGCGVQGNLETYL 65
DB 20 LILLILGVALCSAVEKKVCGGINNKLTQGLGHVBDHFTSLQRMNCEVLSNLEITY 79
QY 66 PTNASTLFLDIOEVGVYVLIANQVROVPLQRLRIYRGTOLEFEDNVALAVLDNGDPLNN 125

Db 80 EHNRLDTFLKTKIOEVAGYVILANMADVIPLEMLQITIRGNVLYDNSFALAVLSNYH-NMK 138
 Oy 126 TTPVNTASPGGVELRLSRLTEILKGLQIQRNPOLCYODTLIMKRIFFKNNOLATLTD 185
 Db 139 TO-----GLLELPKRLSELTLNGVKISNNPKLCNMDTVLMNDIIDTSRK-PLTVLD 189
 Oy 186 -TNRBACHPCSPCKSGKSRCKWSESSDCSLPRTVCAGCA-RCKGPLEPTDCHEQCAAG 243
 Db 190 FASNLSSCKPKCHNCTEDHGWGAGEQNCOTLFRVICAQCSGSRCKRVKPSDCCHNCAG 249
 Oy 244 CTGPKHSDCIACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCYTACPYNY 303
 Db 250 CTGPRSDCLACRFRDADATCKDTCPLVLYNTTYOMVNPCKYSGFATCVRECPHNY 309
 Oy 304 LSTDVSGCTLVCPHNOEVAEDGTORCEKSGPCAR 340
 Db 310 VYTDHGSVCVRSCNTDTEYV-EENGVRKCKKCGDLCSK 345

RESULT 6
 GOHUE
 epidermal growth factor receptor precursor - human
 N:contains: protein-tyrosine kinase (EC 2.7.1.112) erBb
 C:Species: Homo sapiens (man)
 C:Date: 15-Nov-1984 #sequence-revision 27-Nov-1985 #text-change 11-Jun-1999
 C:Accession: A00641; A25772; S30024; A38672; A06642; A43615; A23062; A05281; A60143; A33
 R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y
 tg, P.H.
 Nature 309, 418-425, 1984

A>Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
 A:Reference number: A00641; MUID:84219729
 A:Accession: A00641
 A:Molecule type: mRNA
 A:Residues: 1-1210
 A:Cross-References: EMBL:X00588; NID:g31113; PIDD:CAA25240.1; PID:g757924
 A:Note: the authors translated the codon AAG for residue 540 as Asn
 R:Shih, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
 A>Title: Characterization and sequence of the promoter region of the human epidermal gro
 A:Reference number: A25772; MUID:85270438
 A:Accession: A25772
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-29 <ISH>
 A:Cross-References: GB:M11234; NID:g181981; PIDD:AAA52370.1; PID:g553272
 R:Halay, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ulrich, A.; Waterfield, M.
 Oncogene Res. 1, 375-396, 1987
 A>Title: The human EGF receptor gene: structure of the 110 kb locus and identification
 A:Accession: S30024; MUID:88217333
 A:Accession: S30024
 A:Molecule type: DNA
 A:Residues: 1-29 <HA2>
 A:Cross-References: EMBL:X06370; NID:g31118; PIDD:CAA29668.1; PID:g31119
 R:Halay, J.D.; Waterfield, M.D.
 J. Biol. Chem. 266, 1746-1753, 1991
 A>Title: Contributory effects of de Novo transcription and premature transcript terminat
 A:Reference number: A38672; MUID:91107677
 A:Accession: A38672

A:Molecule type: DNA
 A:Residues: 1-29 <HAL>
 A:Cross-References: GB:M8425; NID:g181977; PIDD:AAA63171.1; PID:g553271
 A:Experimental source: carcinoma cell line A431-7
 R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Mer
 Nature 309, 806-810, 1984
 A>Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
 A:Reference number: A00642; MUID:84245635
 A:Accession: A00642
 A:Molecule type: mRNA
 A:Residues: 'RCMARRA', 150-187, 'KSVIOAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321
 ', 798-799, 'ND', 802-811, 'R', 813-942 <XU>
 A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF recep
 R:Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.

Science 224, 843-848, 1984
 A>Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati
 A:Reference number: A43615; MUID:84196372
 A:Accession: A43615

A:Molecule type: mRNA
 A:Residues: 713-964 <LIN>
 A:Experimental source: epidermoid carcinoma cell line A431
 R:Simmen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984
 A:Reference number: A23062; MUID:85046483
 A:Accession: A23062

A:Molecule type: mRNA
 A:Residues: 1028-1210 <SIM>
 R:Weber, W.; Gyll, G.N.; Speiss, J.
 Science 224, 294-297, 1984
 A:Reference number: A05281; MUID:84172183
 A:Accession: A05281

A:Molecule type: protein
 A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
 R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A>Title: Identification of residues in the nucleotide binding site of the epidermal g
 A:Reference number: A60143; MUID:85182650
 A:Accession: A60143

A:Molecule type: protein
 A:Residues: 740-744, 'X', 746-747 <RUS>
 R:Roczowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A>Title: ATP-stimulated interaction between epidermal growth factor receptor and supe
 A:Reference number: A38023; MUID:84191554
 A:Contents: annotation: receptor activity
 A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C
 Cell 59, 33-43, 1989

A>Title: Functional independence of the epidermal growth factor receptor from a doma
 A:Reference number: A3331; MUID:90003233
 A:Contents: annotation: internalization signal
 C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-recepto
 C:Genetics:
 A:Gene: GDB:EGFR
 A:Cross-References: GDB:120610; OMIM:131550

A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1210/Product: EGF receptor #status predicted <KAT>
 F:25-645/Domain: extracellular #status predicted <EXT>
 F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>
 F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F:646-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:710-975/Domain: protein kinase homology <KIN>
 F:718-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:128, 175, 352, 413, 444, 528, 603/Binding site: carbohydrate (Asn) (covalent) #status pre
 F:745/Active site: Lys #status experimental

Query Match 34.7%; Score 793; DB 1; Length 1210;
 Best Local Similarity 45.3%; Pred. No. 9.7e-47;
 Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

Oy 11 LLLALLPPAA--SYOVCTGDMKRLRSPETHDMLRHLVYOGGOVGOVGNLETLPLTN 68
 Db 14 LLAALCPASRALEEKVKVCGTGNKTLQGLTFEDHPLSLQRMNCEVVALGNLEITVGRN 73
 Oy 69 ASLSFLDIOEVGYVLLAHNOVROVPLQRLRIVGTOLFEDNVALAVLDNGDPLNNTTP 128
 Db 74 YDLSEFLTKIOEVAGYVILANMADVIPLEMLQITIRGNVLYDNSFALAVLSNYH-NMK 126
 Oy 129 VYGASDGLRELQLSLEFLKGLVLIQRNPOLCYODTLIMKRIFFKNNOLATLTDNR 188

Db 127 ---ANKTGKLEPMRNLQETILGAVRFSNNPALCNVESIQMDIVSDPLSNMMDQFNH 183
Qy 189 SRACHPGSMCKGSRCKWGSSESDCSLRTTVACGCA-RCKGPLPTDCCHECAAGCTGP 247
Db 184 LGSCKCKPSCPCNGSCWGGEGENCOKRLIKIAQCCSGCRCKSPSDCHNCACAGCTGP 243
Qy 248 KHSDDIACIHFHSGISCEIHCALVTYNTDPESMPNPEGRTYFASCVTACPYVYLTSD 307
Db 244 RSDPCLCKRFDEATCKDTCPTPLMLYNTTYQMDVNPBGKISFGATCVKCKPRNYVTD 303
Qy 308 VGSCTLVCPPLHNOEVTAEADGTORCEKCSKPCAR 340
Db 304 HGSVACGACADSEYK-EEGCVARKCKCKGPCRK 335
RESULT 7
A53183
epidermal growth factor receptor precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
C:Accession: A53183; A43818; S24942; A28941; S45325; I49643
R:Rutledge, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.;
Genes Dev. 6, 399-413, 1994
A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
A:Reference number: A53183; MUID:94170986
A:Accession: A53183
A:Molecule type: mRNA
A:Residues: 1-1210 <LUE>
A:Cross-references: GB:U03425
R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A:Reference number: A43818; MUID:91232866
A:Accession: A43818
A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Eislinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942
A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA5587.1; PID:G488831
R:Patla, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A:Reference number: I49643; MUID:93126380
A:Accession: I49643
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:I06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: Atp; growth factor receptor; kinase-related transforming protein; phosphoprotein

F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase Atp-binding motif
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental
Query Match 34.5%; Score 789; DB 2; Length 1210;
Best Local Similarity 46.3%; Pred. No. 1,86-46;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;
Qy 11 LLLALLPQAA--STQVCTGDMKRLPASPETHLMDLRHLHYGCGVQVGNLELTPTN 68
Db 14 LITFALCAAGGALAEKKVCGQTSNRLTQGTEDHFLSIQRMVNNCEVYALGNLEITYQRN 73
Qy 69 ASLSPLDIOEVQGVVLLAHNOVROVPLQRLRYNGTQLFEDNVALAVLDNGDPLNNTTP 128
Db 74 YDLSPLKTIQEVAGVLLALNTVERIPLENQIIRGNALYENTYALATLSN----- 124
Qy 129 VTGASPGGLRELOLRSLTEILKGVLLQIRNPOLCYODITLWKDI---FHKRNQALATLI 184
Db 125 -YGTNRITGLRELPMRNLQETILGAVRFSNNPALCNVESIQMDIVSDPLSNMMDQFNH 180
Qy 185 DTRNSRACHPCSPMCKGSRCKWGSSESDCSLRTTVACGCA-RCKGPLPTDCCHECAAG 243
Db 181 -QSHPSCKPCDPPSCPNCSWGGEGENCOKRLIKIAQCCSGCRCKSPSDCHNCACAG 239
Qy 244 CTGPRHSCILACIHFHNSGICELHCPALVYNTDFFESMPNPEGRTYFASCVTACPYVY 303
Db 240 CTGPRSDCLVCKRFQDEATCKDTCPTPLMLYNTTYQMDVNPBGKISFGATCVKCKPRNY 299
Qy 304 LSTDVGSCTLVCPPLHNOEVTAEADGTORCEKCSKPCAR 340
Db 300 VTDHGSVCVRACGADSEYK-EEGCVARKCKCKGPCRK 335
RESULT 8
A36325
epidermal growth factor receptor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 25-Jan-1991 #sequence revision 25-Jan-1991 #text change 10-Oct-1997
C:Accession: A36325
R:Petich, L.A.; Harris, J.; Raymond, V.W.; Blasband, A.; Lee, D.C.; Earp, H.S.
Mol. Cell. Biol. 10, 2973-2982, 1990
A:Title: A truncated, secreted form of the epidermal growth factor receptor is encoded
A:Reference number: A36325; MUID:90258888
A:Accession: A36325
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-644 <PEP>
A:Cross-references: GB:X47394
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; Atp; growth factor receptor
Query Match 34.2%; Score 781.5; DB 2; Length 644;
Best Local Similarity 46.0%; Pred. No. 3,1e-46;
Matches 157; Conservative 44; Mismatches 117; Indels 23; Gaps 7;
Qy 3 LALICRMGLLALLPQAA-STQVCTGDMKRLPASPETHLMDLRHLHYGCGVQVGNLE 61
Db 15 LALCAAG-----GALAEKKVCGQTSNRLTQGTEDHFLSIQRMVNNCEVYALGNLE 66
Qy 62 LTYLPTNASTLSPLDIOEVQGVVLLAHNOVROVPLQRLRYNGTQLFEDNVALAVLDNGD 121
Db 67 ITTYVQRNDLSFLKTIQEVAGVLLALNTVERIPLENQIIRGNALYENTYALAVLSN-- 124
Qy 122 PLNNTTPVTGASPGGLRELOLRSLTEILKGVLLQIRNPOLCYODITLWKDI--FHKRNQAL 181

Db 125 -----YGTNKTGLRELPMRNLQELIGAVRFSNNPILCNMERTLOWRDIY-ODVFLSN 175
QY 182 TLIDTNRNRS-RACHPCSPMKGSRMCESSEDCOSLTRVYACAGCA-RCKGPIPTCOCHQ 239
Db 176 MSMDVQRHLTGCPKCDPSPGNSCWRGECNCKLTKITCAQCCSRCKRSPSDCCNQ 235
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFEPMNPEGRYTFGASCVTAC 299
Db 236 CAAGCTGPRSPOLVCHRRDEATCKDTCPLMLYNTTYQMDVNPBGKYSFGACVKKC 295
QY 300 PNYLSTDVSGCTVCPRLHNOEYTAEDGTQRCCKSPCAR 340
Db 296 PNYVYTDHSGCVARACGPDYEV-EDDGVSKCKCKDCGPKRC 335

RESULT 9

A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culloscu, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PILO>
A:Cross-references: GB:L07868; NID:9337359; PID:AA59446.1; PID:9337360
A>Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match 33.9%; Score 775; DB 2; Length 1308;
Best Local Similarity 45.1%; Pred. No. 1.8e-45;
Matches 157; Conservative 45; Mismatches 124; Indels 22; Gaps 7;

QY 9 WGLLALLPPGAA---STOYCTGTDMLRLPASPTHLMDLRHLYOCQVVOGNIETLY 64
Db 8 WYVWSLLVLAAGTVQPSDSQVAGTENKLSLSLEQQYRALRKYYENCEVVMGNIETLS 67
QY 65 LPTNASTSLDIOEVOGYVLIANQVROVPLQRLRIYVGTQLEFEDNVALAVLDNDPLN 124
Db 68 IEHNRLSLFLRSREYTGVALNQLRYPLNRLIKGTLYEDRYALALFLNTRKDG 127
QY 125 NTPPYTGAASPGGLRELQRLSLTEILKGYLIQRPOLCYQDTILMKDIFHKNNQALATLI 184
Db 128 NF-----GIQELGKLNTEILNGVYVQDNKFLCADITIHQDIIYRNWPSPSLTIV 178
QY 185 DYNRSRACHPCSPMKGSRMCESSEDCOSLTRVYACAGC-ARCKGPIPTDCCHEQCAAG 243
Db 179 STNGSSGCGCRKRSCTG-RCMGPTENHCQTLTRYCAEQCDGCPRYVSDCHRECAAG 237
QY 244 CGPKHSDCLACLFHNSGICELHCPALVTYNTDFEPMNPEGRYTFGASCVTACPYNY 303
Db 236 CGSPKDTDFACMNFENDSACVYQCPQTFVNYTTFQLEHNENAKTYGAFCKKCPHNF 297
QY 304 LSTDVSGCTVCPRLHNOEYTAEDGTQRCCKSPCAR-----GTHSL 346
Db 298 V-VDSSCYRACPRSSKMEY-EENGIMCKPCPTDICRACDGIGTGLM 343

RESULT 10

A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.1) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epiderma
A:Reference number: A36223; MUID:90083234
A:Accession: A36223

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro,
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor
A:Reference number: I59164; MUID:90311312
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:9183990; PID:AAA3579.1; PID:9306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 32.2%; Score 735.5; DB 2; Length 1342;
Best Local Similarity 44.0%; Pred. No. 9.5e-43;
Matches 150; Conservative 49; Mismatches 121; Indels 21; Gaps 9;

QY 10 GLLALLPPGAA--STOYCTGTDMLRLPASPTHLMDLRHLYOCQVVOGNIETLYPT 67
Db 11 GLFSLANGSEVGNQANCPGLNGLSVTGAENQYQTLTKYECGEVVMGNIETLYLVGH 70
QY 68 NASTSLDIOEVOGYVLIANQVROVPLQRLRIYVGTQLEFEDNVALAVLDNDPLNNTT 127
Db 71 NADLSFLQWIREVTGYVIVAMNEFSTLPLRLRVAGTQVADGKPAIFVM-----LANYT 125
QY 128 PYTGAASPGGLRELQRLSLTEILKGYLIQRPOLCYQDTILMKDIFHKNNQALATLITN 187
Db 126 ---NSSHALRQLRLTQLEILISGVYIEKNDKLCMHPTIDMRDVRD---AEIVKD 178
QY 188 RSRACHPCSPMKGSRMCESSEDCOSLTRVYACAGC-ARCKGPIPTDCCHEQCAAGCTG 246
Db 179 NRSRCPCHVCKG-RCMGPPSEDCQTLTKITCARQCGHCGRPNCCCHDECAAGSG 237
QY 247 PKHSDCLACLFHNSGICELHCPALVTYNTDFEPMNPEGRYTFGASCVTACPYNYLST 306
Db 238 PDQTCFACRHNENDGACVPRCPQLVYVYKLTFLQLEPDPHRTKYQYGVAVASCPHNFV-V 296
QY 307 DVGSGCTVCPRLHNOEYTAEDGTQRCCKSPCAR---GTHS 344
Db 297 DQTSVCVRACPPDKMEVD-KNGLMKCEPCGGLCPKACEGTGS 336

RESULT 11

JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protel
A:Reference number: JC4387; MUID:96096535
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-1339 <REL>
A:Cross-references: GB:U29339; NID:9915389; PID:9915390
A:Experimental source: liver
A>Note: The authors translated the codon AAC for residue 369 as Thr and GTP for resid

C:Comment: This protein is a functional heregulin receptor that transduces signals to tyrosine kinase.

C:Gene: ErbB3

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology

C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1338/Product: epidermal growth factor homolog #status predicted <MAT>

F:640-659/Domain: transmembrane #status predicted <TM>

F:705-970/Domain: protein kinase homology <KIN>

F:713-721/Region: protein kinase ATP-binding motif

F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (C

Query Match 30.3%; Score 692; DB 2; Length 1339;

Best Local Similarity 41.8%; Pred. No. 9.1e-40;

Matches 146; Conservative 46; Mismatches 131; Indels 26; Gaps 10;

3 LAALRQGLALLLPFGA---STQVCTGDMKLLPASPENHDMLEHLYOGCOVQGN 59

7 LQVLC---FLSLIARGSEMGNSQAVCPPTGLSVTDADADQVTLKLYEKCVQMGN 62

60 LEIATLPFNASLSFQDIOEVOGYVLIHNGVRQVPLRLRVSTGLFEDNYALAVDN 119

63 LEIVLTGNADSLFQMLREVTAYVLYVAMNERSVLPRLNLRVSTQYDDGFAFVA-- 120

120 GDPNNTPPVYTCASPGSLREQLRSLEILKGVLIQRNPOLCYQDTILMKDIFKNNQL 179

121 ---LNYNT---NSSHALROLKFTQLETLLSGVYIEKNDKLCIHMDTIDMRDLYVR-- 170

180 ALTIIDTRNSRACHCSPMKGSRWGSESSDDQSLTFTVCAAGC-ARCKPLPLPDCHE 238

171 GAELVYKKNAGCPCPCHEVCKG-RCMGRPPDCCQLITRTICAPQCNCRGCFGNPQCCHD 229

239 QCAAGCTCPKHSDDLACLFHNHSGICELHCPALVTYNTDFESMNPGRGTFGASCYA 298

230 ECAGGCGSGPQDTDCACACRFNDGACVPRCPPLVYNKLLFQLEPNPHTKIYGGCVAS 289

299 CPYNTLSTDVSGCTLVCPRLHNOEVTAEQTORCEKCSKPCAR--GTHS 344

290 CPNFEV-VDQFECVRACPPDKMEVD-KHGLKMCPCGGLCPRACGTS 336

RESULT 12

S06142

Protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish

Alternate names: epidermal growth factor receptor homolog; kinase-related transforming

C:Species: Xiphophorus maculatus (southern platyfish)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000

C:Accession: S06142; S13809

R:Mitbroadt, J.; Adam, D.; Malitschek, B.; Maeneler, W.; Raulf, F.; Telling, A.; Roberts

Nature 341, 415-421, 1989

A>Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu loc

A:Reference number: S06142; MUID:90015140

A:Accession: S06142

A:Molecule type: DNA

A:Residues: 1-1166 <MT>

A:Cross-references: EMBL:X16891; NID:965290; PIDN:CAA34770.1; PID:965291

R:Adam, D.; Maeneler, W.; Scharl, M.

Oncogene 6, 73-80, 1991

A>Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophoru

A:Reference number: S13807; MUID:91125882

A:Accession: S13809

A:Status: preliminary; translation not shown

A:Molecule type: DNA

A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>

A:Cross-references: EMBL:X56319; NID:965284; PIDN:CAA39763.1; PID:965285

C:Gene: mrk

C:Gene: mrk

A:Map position: Y

A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyrosi

F:1-55/Domain: signal sequence #status predicted <SIG>

F/26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F/707-972/Domain: protein kinase homology <KIN>
F/715-723/Region: protein kinase ATP-binding motif

```

Query Match          29.8%; Score 681.5; DB 1; Length 1166;
Best Local Similarity 42.4%; Pred. No. 4,1e-39;
Matches 145; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

OY 4 AALCRMGILLALLPPGAASL-----OVTGTDMKRLRPASPEFHLDMLRHLVGGCOVGN 59
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 8 AALLQ--LLLVISIRCSCTDPDRKVCQGTSNQM---LDNHILKMKKMTSGCNVLEN 62
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
OY 60 LELTYLPTNASLSFLDIOEVOGYVLLAHNOVROYPLRLRIYVGTQLFEDNYALAVLN 119
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 63 LEITFTQENODLSFIQSIOEVOGYVLLAHNEVSTPLVNLRIIGONILEGNEFTLWASN 122
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
OY 120 GDPLNNTPTVYGASGGGRLQLRSLEFLKGVLIQNNPOLCYODTLIMKDFHKNNQL 179
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 123 YOK -NPSS--DVYQVGLKQLDLSNLTETLSGGVVSINPLLCNVTIMMDIVDKTSNP 179
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
OY 180 ALTLIDITNRSRACHCSPKCKSGCSESESDCOSLFTVYCAAGC -ARCKGPLPTDCHE 238
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 180 TYNLILPHAFERCCQCKDHGCVAGSGCAPRGHCQKFTLLIDAECCNRRCRPPKIDCNE 239
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
OY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDFESMPNPGRYTFGASCVTA 298
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 240 HCAGGCTGPRATDCLACRDFNDGTCKDTPCPKPYLDIVSHQVVDNPIKKTFCGACVKE 299
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
OY 299 CPYNTLSIDVGSCTLVCPRLHNOEVTAEQGTQCEKCSFPCAR 340
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 300 CPSNYVYTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPK 339
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 13
A27131
epidermal growth factor receptor - fruit fly (Drosophila melanogaster) (fragment)
C:Species: Drosophila melanogaster
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 23-May-1997
C:Accession: A27131
R:Schaejter, E.D., Segal, D.; Glazer, L.; Shilo, B.Z.
Cell 46, 1091-1101, 1986
A:Title: Alternative 5' exons and tissue-specific expression of the Drosophila EGF re
A:Reference number: A27131; MUID:87002474
A:Accession: A27131
A:Molecule type: mRNA
A:Residues: 1-843 <SCH>
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor

Query Match          25.2%; Score 575.5; DB 2; Length 843;
Best Local Similarity 36.6%; Pred. No. 5.4e-32;
Matches 119; Conservative 45; Mismatches 130; Indels 31; Gaps 7;

OY 24 QVCTGDMKRLRPASPEFHLDMLRHLVGGCOVGNLELYLPT-MSLSFLDIOEVOG 82
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 51 KVCIGTKSRLSYKSHENHRYRLRDYNTCYVDGNLKLTMLPENIDLSFLDNIREVTTG 110
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
OY 83 YVLIAHNOVROYPLRLRIYVGTQLF-----EDNTALAVIDNGRPLNNTTPTVYGASRGL 137
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 111 YVLIASHVDKRVKVPRLDIKGRITLSLVEBEKALTV-----TYSKM 154
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
OY 138 RELQRLSLTEILKGVLIQNNPOLCYODTLIMKDFHKNNOLATLTLIDITNRSRACHPCSP 197
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 155 YTLLEPLRLDRLVANGVGHNNYNTLCHMRTIQMSVLSNGTAYIYNDPTAAERECRCKHE 214
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
OY 196 MCKSGRCWGESSEDCQSLRTFYVACAGCA -RCKGPLPTDCCHEDCAAGCTGPKHSDCLAC 255
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 215 SCTTH-CWEGEGRKNQKFSKTLTSCSPQAGRCGRYGRPKPECCNHLFCAGAGCTPTGDKDIAC 273
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```

OY 256 LFNHSGICEHCPALVYNTDTFESMPNREGRTYEGASCATAPYNYLTDVSGCTLYC 315
DB 274 KNFEDVANSKECCPPKRYNPTTYLETINBGRKAIYGAICVKECP-CHLLRDNGACVNSC 332
OY 316 PLHNOEVTAEADGTORCEKCSKPCAR 340
DB 333 PQDKMDKGE-----CVPNGPCPK 352
RESULT 14
S70712
protein-tyrosine kinase (EC 2.7.1.112) let-23 precursor - Caenorhabditis elegans
N.Alternate names: receptor tyrosine kinase let-23
C.Species: Caenorhabditis elegans
C.Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 17-Nov-2000
C.Accession: S70712; S73101; S13422; T27682
R.Sakai, T.; Koga, M.; Ohshima, Y.
J.Mol. Biol. 256, 548-555, 1996
A.Title: Genomic structure and 5' regulatory regions of the let-23 gene in the nematode
A.Reference number: S70712; MUID:96177760
A.Accession: S70712
A.Status: nucleic acid sequence not shown
A.Molecule type: DNA
A.Residues: 1-1374 <SAK>
A.Cross-references: EMBL:D63426
A.Experimental source: strain N2
R.Koga, M.
submitted to the EMBL Data Library, July 1995
A.Reference number: S73101
A.Molecule type: DNA
A.Accession: S73101
A.Residues: 1-50, 'G', 52-1374 <KOG>
A.Cross-references: EMBL:D63426; NID:g1407562; PIDN:BA009729.1; PID:d1010375; PID:g14075
R.Ariolan, R.V.; Koga, M.; Mendel, J.E.; Ohshima, Y.; Sternberg, P.W.
Nature 348, 693-699, 1990
A.Title: The let-23 gene necessary for Caenorhabditis elegans vulval induction encodes a
A.Reference number: S13422; MUID:91080919
A.Accession: S13422
A.Molecule type: mRNA
A.Residues: 52-1374 <ARO>
R.Thomas, K.
submitted to the EMBL Data Library, March 1996
A.Reference number: 220404
A.Accession: T27682
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 52-1374 <MIL>
A.Cross-references: EMBL:270038; PIDN:CA93882.1; GSPDB:GNO0020; CESP:ZK1067.1
A.Experimental source: clone ZK1067
C.Genetics:
A.Gene: let-23; CESP:ZK1067.1
A.Map position: 2
A.Introns: 44/1; 51/1; 90/1; 112/3; 165/3; 217/1; 290/1; 379/1; 418/1; 448/2; 565/1; 608
C.Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C.Keywords: ATP; phosphotransferase; transmembrane protein; tyrosine-specific protein ki
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1374/Product: protein-tyrosine kinase let-23 #status predicted <MAT>
F:934-1199/Domain: protein kinase homology <KIN>
F:942-950/Region: protein kinase ATP-binding motif

Query Match 18.3%; Score 419; DB 2; Length 1374;
Best Local Similarity 29.0%; Pred. No. 4.8e-21;
Matches 102; Conservative 59; Mismatches 131; Indels 60; Gaps 14;

OY 25 VCTGDMKLRLPASPEHLDMLRHLRYOCQOVQGNLELTYPN----- 68
DB 90 LCGTNGISRGCTGNI-LEDEFTWYRGCRRYGNGLEITWIANELKKRSTNSTVDPK 148
OY 69 -----ASLSTQDIQEOVGVLIAHNOVQVPLQRLRIYRGTOLEEDNYALAVLDNGDP 122
DB 363 HCVKECPPELLIXND-----VCVRHCSGHHYDATKDMRECEKCPSSSC-----PKI 409

DB 149 NEDSEPKSINFENLEIRGSLIIYRANIKSPRLRVIGDEVEFHDN-ALYIHKNDK- 206
OY 123 LNNTPRYGASPGGLREQLSLTEFLKGVLIQRNPQLC-QDTILMKDLFHHNNOLAL 181
DB 207 -----VHEVYMRRLRYIRNGSVTIDNPMCYIGKIDMKELLYDPD-VQ 250
OY 182 TLIDTNRSRACH-----PCSPMKSGRCGSESEDCOSLTRVYACAGCARC---KGPL 231
DB 251 KYETTNSHOCYQNKCKSMNAKCHESC-NDKCWGSGNDQORYRYSVCPSKSCQCFYSNSTS 309
OY 232 PTDCEHCQACGCTGPKHSDCLAHFNHSGICEHCPALVYNTDTFESMPNREGRTY 291
DB 310 SYECDSDACLGCTGHHGPKNCIACSKYELDICTECPSRKIFNFKTGRLEFNDGRYON 369
OY 292 GASCYTACPYNTL-STDVGSLTVC-PLHNOEVTAEADGTORCEK-SKPCAR 340
DB 370 GNCVKECPPELLIENDV--CVRRHCSGHHYDATKDMRECEKCRSSSCP 417
RESULT 15
S70713
protein-tyrosine kinase let-23 precursor homolog - Caenorhabditis vulgaris
N.Alternate names: receptor tyrosine kinase let-23 homolog
C.Species: Caenorhabditis vulgaris
C.Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Dec-1997
R.Sakai, T.; Koga, M.; Ohshima, Y.
J.Mol. Biol. 256, 548-555, 1996
A.Title: Genomic structure and 5' regulatory regions of the let-23 gene in the nemato
A.Reference number: S70712; MUID:96177760
A.Accession: S70713
A.Status: nucleic acid sequence not shown
A.Molecule type: DNA
A.Residues: 1-1369 <SAK>
A.Cross-references: EMBL:D63427
C.Genetics:
A.Gene: let-23
A.Introns: 42/1; 49/1; 83/1; 105/3; 155/3; 207/1; 280/1; 369/1; 408/1; 438/2; 555/1;
C.Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C.Keywords: ATP; phosphotransferase; transmembrane protein; tyrosine-specific protein
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1369/Product: protein-tyrosine kinase let-23 homolog #status predicted <MAT>
F:929-1194/Domain: protein kinase homology <KIN>
F:937-945/Region: protein kinase ATP-binding motif

Query Match 18.2%; Score 416; DB 2; Length 1369;
Best Local Similarity 28.6%; Pred. No. 7.7e-21;
Matches 105; Conservative 59; Mismatches 135; Indels 68; Gaps 13;

OY 25 VCTGDMKLRLPASPEHLDMLRHLRYOCQOVQGNLELTYPN----- 68
DB 83 VCSGNNILSRGSGNT-LEDEHMYRRCRRYGLLETTWIANELQKWRSTNQTVDAD 141
OY 69 -----ASLSEFLDIOEVOGVYLIHNOVQVPLQRLRIYRGTOLEEDNYALAVLDNGDPLN 124
DB 142 IDYLKTVNFEHLERISLIYRANIKISPKRLRVIGDEVFHDN-SLYIHQHEK--- 197
OY 125 NTTPYTGASPGGLRLQRLSLTEFLKGVLIQRNPQLCYOT-ILMKIPIKKNQDLATL 183
DB 198 -----VNLVWKLELRVINGSVSIONNPMFLTKDWMNILLDXSRQVE- 244
OY 184 IDTNRSRACHPCSPM-----CKGRCMGSESEDCOSLTRVYACAGCARCKGPLPT--- 233
DB 245 -XXNSHKACWANGELIAXKHENCK-DKCGWGDNDQCYIYSVCPKPSQCFYSITDYSY 302
OY 234 DCEHDCQACGCTGPKHSDCLAHFNHSGICEHCPALVYNTDTFESMPNREGRTYFGA 293
DB 303 EECDSGSLGCGCNIHNRDSCIACSKYEMDMCIDCPARKKINHHKTRGLRVPRDGRYQNGN 362
OY 294 SCVTACPYNTYSTDVGSLTVCPLHNOEVTAEADGT---ORCEK-SKPCARHTSLRPR 349
DB 363 HCVKECPPELLIXND-----VCVRHCSGHHYDATKDMRECEKCPSSSC-----PKI 409

OY 350 AAVVPL 356
| |
Db 410 CTVDGPL 416

Search completed: April 11, 2002, 09:30:11
Job time: 108 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:34:18 ; Search time 15.52 Seconds

(Without alignments)
989.856 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287

Sequence: 1 MELALCRMGILLALLPFGA.....VGKGPDAHVAVNLRYEG 419

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	1878	82.1	1255	1	ERB2_HUMAN
2	1608.5	70.3	1257	1	ERB2_RAT
3	1571	68.7	1254	1	ERB2_MESAU
4	796.5	34.8	703	1	EGFR_CHICK
5	793	34.7	1210	1	EGFR_HUMAN
6	789	34.5	1210	1	EGFR_MOUSE
7	775	33.9	1308	1	ERB4_HUMAN
8	735.5	32.2	1342	1	ERB3_HUMAN
9	698	30.5	1339	1	ERB3_RAT
10	681.5	29.8	1167	1	XMRK_XIPMA
11	574.5	25.1	1426	1	EGFR_DROME
12	419	18.3	1323	1	LT23_CAEEL
13	342.5	15.0	1363	1	ILPR_BRALA
14	294	12.9	2146	1	INSR_DROME
15	291	12.7	1477	1	HTK7_HYDAT
16	278	12.2	1300	1	IRR_CAVPO
17	270	11.8	581	1	IRR_RAT
18	269.5	11.8	1607	1	MITR_LYMST
19	269	11.8	1297	1	IRR_HUMAN
20	264.5	11.6	1382	1	INSR_HUMAN
21	263.5	11.5	1383	1	INSR_RAT
22	261	11.4	1372	1	INSR_MOUSE
23	258	11.3	1390	1	INSR_AEDAE
24	257.5	11.3	1367	1	IGIR_HUMAN
25	252.5	11.0	1370	1	IGIR_RAT
26	250.5	11.0	1373	1	IGIR_MOUSE
27	169	7.4	1696	1	PCK5_BRACL
28	144	6.3	1877	1	PCK5_MOUSE
29	142	6.2	1959	1	AGRI_RAT
30	131.5	5.7	913	1	PCK5_HUMAN
31	129.5	5.7	1680	1	FUR2_DROME
32	124.5	5.4	1877	1	PCK5_RAT
33	123.5	5.4	937	1	PAC4_RAT

34	123	5.4	417	1	WSL1_HUMAN	Q93038 h wsl-1 pro
35	119.5	5.2	667	1	TS11_GIALA	Q03185 giardia lam
36	119	5.2	2248	1	ZAN_RABIT	P57999 oryctolagus
37	118.5	5.2	551	1	LEM2_RABIT	P27113 oryctolagus
38	118.5	5.2	3635	1	LMAS_MOUSE	Q61001 mus musculu
39	117	5.1	469	1	PROP_HUMAN	P27918 homo sapien
40	117	5.1	3075	1	LMAL_HUMAN	P25391 homo sapien
41	116.5	5.1	969	1	PAC4_HUMAN	P29132 homo sapien
42	116.5	5.1	1173	1	TSP1_XENLA	P35448 xenopus lae
43	115	5.0	3106	1	LMAS_MOUSE	Q60675 mus musculu
44	114.5	5.0	484	1	LEM2_PIG	P98110 sus scrofa
45	113	4.9	1122	1	TIE2_MOUSE	Q02858 mus musculu

ALIGNMENTS

RESULT	ID	ERB2_HUMAN	STANDARD	PRT	1255 AA.
1	AC	P04626	13-AUG-1987 (Rel. 05, Created)		
	DT	13-AUG-1987 (Rel. 05, Last sequence update)			
	DT	20-AUG-2001 (Rel. 40, Last annotation update)			
	DE	RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2 PRECURSOR (EC 2.7.1.112)			
	DE	(P185ERBB2) (NEU PROTO-ONCOGENE) (C-ERBB-2) (TYROSINE KINASE-TYPE CELL SURFACE RECEPTOR HER2) (MLN 19).			
	GN	ERBB2 OR HER2 OR NGL OR NEU.			
	OS	Homo sapiens (Human).			
	OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	OX	NCBI_TaxID=9606;			
	RN	[1]			
	RP	SEQUENCE FROM N.A.			
	RX	MEDLINE=86118663; PubMed=3003577;			
	RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,			
	RA	Saito T., Toyoshima K.;			
	RT	"Similarity of protein encoded by the human c-erb-B-2 gene to			
	RT	epidermal growth factor receptor.";			
	RL	Nature 319:230-234(1986).			
	RN	[2]			
	RP	SEQUENCE FROM N.A.			
	RX	MEDLINE=86070181; PubMed=2999974;			
	RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,			
	RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,			
	RA	France U., Levinson A., Ullrich A.;			
	RT	"Tyrosine kinase receptor with extensive homology to EGF receptor			
	RT	shares chromosomal location with neu oncogene.";			
	RL	Science 230:1132-1139(1985).			
	RN	[3]			
	RP	SEQUENCE OF 737-1031 FROM N.A.			
	RX	MEDLINE=86016729; PubMed=2995967;			
	RA	Semba K., Kanata N., Toyoshima K., Yamamoto T.;			
	RA	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the			
	RT	c-erbB-1/epidermal growth factor receptor gene and is amplified in a			
	RT	human salivary gland adenocarcinoma.";			
	RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).			
	RN	[4]			
	RP	VARIANTS VAL-654 AND VAL-655.			
	RX	MEDLINE=93194196; PubMed=8095488;			
	RA	Ehman A., Low J., Wallace R.B., Wu A.M.;			
	RT	"Characterization of a new allele of the human ERBB2 gene by allele-			
	RT	specific competition hybridization.";			
	RL	Genomics 15:426-429(1993).			
	CC	-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
	CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
	CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
	CC	ALPHA AND AMPHIREGULIN.			
	CC	-1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +			
	CC	PROTEIN TYROSINE PHOSPHATE.			
	CC	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
	CC	(POTENTIAL).			
	CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11767; AAA35808.1; JOINED.
CC EMBL; M11761; AAA35808.1; JOINED.
CC EMBL; M11762; AAA35808.1; JOINED.
CC EMBL; M11763; AAA35808.1; JOINED.
CC EMBL; M11764; AAA35808.1; JOINED.
CC EMBL; M11765; AAA35808.1; JOINED.
CC EMBL; M11766; AAA35808.1; JOINED.
CC EMBL; M11730; AAA35493.1; JOINED.
CC EMBL; M12036; AAA35978.1; JOINED.
CC EMBL; X03363; CAA27060.1; JOINED.
CC PIR; A25491; A25491.
CC PIR; A24571; A24571.
CC HSSP; P11362; IFEI.
CC MIM; 164870; -.
CC InterPro: IPR000494; EGFR_L.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_kin.
CC Pfam; PF00757; Furin-like.
CC Pfam; PF01030; Recept_L_domain; 2.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 1255
FT DOMAIN 22 652
FT TRANSMEM 653 675
FT DOMAIN 720 987
FT NP_BIND 726 734
FT BINDING 753 753
FT ACT_SITE 845 845
FT MOD_RES 1139 1139
FT MOD_RES 1248 1248
FT CARBOHYD 68 68
FT CARBOHYD 124 124
FT CARBOHYD 187 187
FT CARBOHYD 259 259
FT CARBOHYD 530 530
FT CARBOHYD 571 571
FT CARBOHYD 629 629
FT VARIANT 654 654
FT VARIANT 655 655
FT VARIANT 655 655
FT CONFLICT 1170 1170
FT SEQUENCE 1255 AA; 137909 MW; 39B9DFDA04DC6F62 CRC64;
Query Match 82.1%; Score 1878; DB 1; Length 1255;
Best Local Similarity 83.0%; Pred. No. 3e-139;

Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
QY 1 MEIATLCRWGILLALIPGASTOYCTGTDMLRLPSPFHLMRLHYGCGVGNL 60
DB 1 MELALCLRWGILLALIPGASTOYCTGTDMLRLPSPFHLMRLHYGCGVGNL 60
QY 61 ELTYLPYASLSFLODIOGVGYVLIANOVROYPLQRLRVGTOLFEEDNYALAVLNDG 120
DB 61 ELTYLPYASLSFLODIOGVGYVLIANOVROYPLQRLRVGTOLFEEDNYALAVLNDG 120
QY 121 DPLNTPYVPGASPGLELEQLRLSTLEIKGCVLIQRNPOLCYDPTILMKDIFRKNOLA 180
DB 121 DPLNTPYVPGASPGLELEQLRLSTLEIKGCVLIQRNPOLCYDPTILMKDIFRKNOLA 180
QY 181 LTLIDNRSRCHCPSPCKSKRCMGESSEPCOSTLRVCGAGCARCGPLPTCCHEOC 240
DB 181 LTLIDNRSRCHCPSPCKSKRCMGESSEPCOSTLRVCGAGCARCGPLPTCCHEOC 240
QY 241 AAGCTGPKHSDCLAFHNSGICELHCPALVTYNTDFFESMPNDEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLAFHNSGICELHCPALVTYNTDFFESMPNDEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPHNOEVTAEADGTQRCCKSPCAR-----GTHSLPRAANVPV 355
DB 301 YNYLSTDVSGCTLVCPHNOEVTAEADGTQRCCKSPCARCYGIGMHLREVRVAVTSAN 360
QY 356 LRMQPG--PAHPVLSFLRPSMDIVSAFSLPLAPLSPISVPI-----SPVSGRGPD 405
DB 361 IQERAGCKKIRGSLAFLESDPDGPAWNT---APLQEPQLOVFETLEITGYLYTSAMPD 417
QY 406 --PDAHVAVNLRYEG 419
DB 418 SLPDLSTVFQNLQVIRG 433
RESULT 2
ID ERB2_RAT STANDARD; PRT; 1257 AA.
AC P06494.
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ERB2-2 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
DE (P185ERB2) (NEU PROTO-ONCOGENE) (EPIDERMAL GROWTH FACTOR RECEPTOR-RELATED PROTEIN).
GN ERB2 OR NEU
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related protein".
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system".
RL Neuron 6:691-704(1991).
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofis F.J., Doak D.G., Mulvey D., Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein.";

RL EMBL J. 11:43-48(1992).
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X03362; CAA27059.1; ALT_INT.
 DR PIR: A24562; TVRTNU.
 DR HSSP: P11362; 1FG1.
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 1257
 FT DOMAIN 22 654
 FT TRANSMEM 655 677
 FT DOMAIN 678 1257
 FT DOMAIN 159 369
 FT DOMAIN 473 646
 FT DOMAIN 722 989
 FT NP_BIND 728 736
 FT BINDING 755 755
 FT ACT_SITE 847 847
 FT MOD_RES 1141 1141
 FT MOD_RES 1250 1250
 FT CARBOHYD 68 68
 FT CARBOHYD 188 188
 FT CARBOHYD 260 260
 FT CARBOHYD 532 532
 FT CARBOHYD 573 573
 FT CARBOHYD 631 631
 FT VARIANT 661 661
 FT SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;
 Query Match 70.3%; Score 1608.5; DB 1; Length 1257;
 Best Local Similarity 85.0%; Pred. No. 36e-118;
 Matches 294; Conservative 15; Mismatches 36; Indels 1; Gaps 1;
 QY 1 MELALACRMGLIALLPFGASTOVCTGDMKRLPASPETHLMDLRHLVYGCQVVGQNL 60
 DB 1 MELAMCMRGFLALPLPGIAGTGYCTGDMKRLPASETHLMDLRHLVYGCQVVGQNL 60
 QY 61 ELTYLPANASLSFLDIOEVOGYMLIAHQVAVPLQRLIRIVGTQLEEDNALAVLDNG 120
 DB 61 ELTYLPANASLSFLDIOEVOGYMLIAHQVAVPLQRLIRIVGTQLEEDNALAVLDNG 120

DB 61 ELTYLPANASLSFLDIOEVOGYMLIAHQVAVPLQRLIRIVGTQLEEDNALAVLDNG 120
 QY 121 DLINNTTPT-GASGGURELOLSLTLLKGGVLIQRPOLCYODTIMKDFPKNNOL 179
 DB 121 DLINNTTPTGASGGURELOLSLTLLKGGVLIQRPOLCYODTIMKDFPKNNOL 180
 QY 180 ALTLIDNRSRACHSPCKSGKSGESSESDCQLTFTVACGAGCARCKPPTCCHEQ 239
 DB 181 AVVDIDNRSRACHSPCKSKNNHMGESPECCQLTGTICSGCARCKRLPTCCHEQ 240
 QY 240 CAGCTGPKHSDCLACLHNSHGICELHCPALVYNTDTFESMPNDEGRYTFGASCVTAC 299
 DB 241 CAGCTGPKHSDCLACLHNSHGICELHCPALVYNTDTFESMNPDEGRYTFGASCVTAC 300
 QY 300 PNYNLTSDVGSCTVCPILNDEVTADGROREKSKPCARTSHL 345
 DB 301 PNYNLTSDVGSCTVCPILNDEVTADGROREKSKPCARTSHL 346
 RESULT 3
 ID ERB2_MESAU STANDARD; PRT; 1254 AA.
 AC Q60553;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ERBB-2 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
 DE (P185ERBB2) (NEU PROTO-ONCOGENE).
 GN ERBB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 CC Mesocricetus.
 OC Neocricetus.
 RX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94193007; Pubmed=7908275;
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RL Gene 140:251-255(1994).
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D16295; BAA03801.1; -
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; Tyrc; 1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane, Glycoprotein, Multigene family; Receptor; Signal.
 KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 1254
 FT TRANSMEM 22 652
 FT TRANSMEM 653 675
 FT DOMAIN 676 1254
 FT DOMAIN 158 368
 FT DOMAIN 472 644
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT MOD_RES 1139 1139
 FT MOD_RES 1247 1247
 FT CARBOHYD 68 68
 FT CARBOHYD 125 125
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT VARIANT 658 658
 FT VARIANT 659 659
 FT SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2B81 CRC64;

Query Match 68.7%; Score 1571; DB 1; Length 1254;
 Best Local Similarity 74.4%; Pred. No. 3 1e-115;
 Matches 299; Conservative 26; Mismatches 67; Indels 10; Gaps 3;

QY 1 MELALCRWGLLALLPGAASTVCTGDMKRLPASETHLDMRLHYGCGVVGML 60
 DB 1 MELALCRWGLLALLPGAASTVCTGDMKRLPASETHLDMRLHYGCGVVGML 60
 QY 61 ELTYLPNTNLSFTDIOIOVGVLIAHNOVQVPLQRIYRGQLFEDNALANLVDNG 120
 DB 61 ELTYLPNTNLSFTDIOIOVGVLIAHNOVQVPLQRIYRGQLFEDNALANLVDNR 120
 QY 121 DPLNNTPTVGTASPGGLRELQLSLTEILKGVLIQIORNQLCYODITLKKDIFKKNQIA 180
 DB 121 DPLNNTPTVGTASPGGLRELQLSLTEILKGVLIQIORNQLCYODITLKKDIFKKNQIA 180
 QY 121 DPLNNTPTVGTASPGGLRELQLSLTEILKGVLIQIORNQLCYODITLKKDIFKKNQIA 180
 DB 121 DPLNNTPTVGTASPGGLRELQLSLTEILKGVLIQIORNQLCYODITLKKDIFKKNQIA 180
 QY 181 LTLIDTNRSPACHPCSPCKSGSRWGESSEDCOSLTRIVCAGGCARCKGPLPTDCHEQC 240
 DB 181 LTLIDTNRSPACHPCSPCKSGSRWGESSEDCOSLTRIVCAGGCARCKGPLPTDCHEQC 240
 QY 241 AAGCTGPKHSDDLACHFNHSGICELHCPALVYNTDFESMPNPGRTVFGASCTTAP 300
 DB 241 AAGCTGPKHSDDLACHFNHSGICELHCPALVYNTDFESMPNPGRTVFGASCTTAP 300
 QY 241 AAGCTGPKHSDDLACHFNHSGICELHCPALVYNTDFESMPNPGRTVFGASCTTAP 300
 DB 241 AAGCTGPKHSDDLACHFNHSGICELHCPALVYNTDFESMPNPGRTVFGASCTTAP 300
 QY 301 YVNLSTVGSCTLVCPILHNOEYTAEDGTORCEKSKPCAR----GTHSLPRPAVPVP 355
 DB 301 YVNLSTVGSCTLVCPILHNOEYTAEDGTORCEKSKPCAR----GTHSLPRPAVPVP 355
 QY 301 YVNLSTVGSCTLVCPILHNOEYTAEDGTORCEKSKPCAR----GTHSLPRPAVPVP 355
 DB 301 YVNLSTVGSCTLVCPILHNOEYTAEDGTORCEKSKPCAR----GTHSLPRPAVPVP 355
 QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLLPLAPISPTSVPI 395
 DB 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLLPLAPISPTSVPI 395
 QY 361 IOEFGACKKIFGSLAFIPESFD---GNPSSGIAPLTPPEQLQV 399
 DB 361 IOEFGACKKIFGSLAFIPESFD---GNPSSGIAPLTPPEQLQV 399

RESULT 4
 EGF_R_CHICK
 ID EGF_R_CHICK STANDARD; PRT; 703 AA.

DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112) (CER
 DE (FRAGMENT)).

GN EGF.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88261272; Pubmed=3260329;
 RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
 RT Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
 RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
 RT expression in mouse cells, and differential binding of EGF and
 RT transforming growth factor alpha."
 RL Mol. Cell. Biol. 8:1970-1978(1988).
 CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 CC AND ALSO OF TGF-ALPHA, AMPHIRGULIN, HEPARIN-BINDING EGF, GP30 AND
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- MISCELLANEOUS: BINDING OF EGF TO THE RECEPTOR LEADS TO
 CC INTERNALIZATION OF THE EGF-RECEPTOR COMPLEX, INDUCTION OF THE
 CC TYROSINE KINASE ACTIVITY, STIMULATION OF CELL DNA SYNTHESIS, AND
 CC CELL PROLIFERATION.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: M20386; AAA48760.1; -.
 DR Interpro: IPR000494; EGF_R_L.
 DR Interpro: IPR000719; Euk_pkinase.
 DR Interpro: IPR002174; Furin-like.
 DR Interpro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 4.
 DR DR PROSITE: PS00107; PROTEIN_KINASE_ATP; PARTIAL.
 DR DR PROSITE: PS00109; PROTEIN_KINASE_TYR; PARTIAL.
 DR DR PROSITE: PS50011; PROTEIN_KINASE_DOM; PARTIAL.
 KM Transmembrane, Glycoprotein, Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 30
 FT CHAIN 31 703
 FT DOMAIN 31 654
 FT TRANSMEM 655 667
 FT TRANSMEM 668 703
 FT DOMAIN 134 134
 FT CARBOHYD 190 190
 FT CARBOHYD 200 200
 FT CARBOHYD 359 359
 FT CARBOHYD 368 368
 FT CARBOHYD 420 420
 FT CARBOHYD 573 573
 FT CARBOHYD 578 578
 FT CARBOHYD 613 613
 FT CARBOHYD 633 633
 FT CARBOHYD 648 648
 FT NON_TER 703 703
 FT SEQUENCE 703 AA; 77427 MW; AEF2DE1B735A690 CRC64;

Query Match 34.8%; Score 796.5; DB 1; Length 703;
 Best Local Similarity 45.5%; Pred. No. 6 3e-55;
 Matches 155; Conservative 55; Mismatches 112; Indels 19; Gaps 7;
 QY 8 RWGLLALLPGAA-----STVCTGDMKRLPASETHLDMRLHYGCGVVGML 61


```

Db      13 RGAAYLVLLLLGVALCSAVEERKVCQGTNNKLTQLGHEVDHFTSLQRMNCEVLSNLE 72
Qy      62 LTYLPTNMSLSTLQDIOEQGVYLAHNOYRVPRLRIYVGTOLFENYLAVALVNDG 121
Db      73 ILYVHNHDLFTLKTIOEVAGYVLLALNWDVPLENIKGNVLYDVSFAVLSNTH 132
Qy      122 PLNNPTVYTGASGGLRELTQLRSLTEILKGVLIORNPOLCQDPTILMKDIFHKNNOLAL 181
Db      133 -MNRKQ-----GLRELPMKRLSEILLNGVKISNNRKLCAMDVYLVNDDITDSRK -PL 182
Qy      182 TLID-TNNSRACHPCSPMCKSGRSGESSEDOQLSTRVYACGCA-RCKGPLPTDCHEQ 239
Db      183 TVLDFASNLSSCPKCHPNCTEDHCWAGEQNCQTLTKVYCAQCGSGRCRGKVPSCCHNQ 242
Qy      240 CAAAGCTGPKHSHCLACLFHNSGICELHCPALVYNTDTFESMPREGRYTTGACVYAC 299
Db      243 CAAAGCTGPRESCCLACRFRDADATCKDPCPLVLYNPPTTYOMDVNPEGRYSFGATCVRBC 302
Qy      300 PYNVLTSDVGSCTVCPPLHNOEVTAEDEGTQRCCKSCPKCAR 340
Db      303 PYNVYVTDHSGCVSRSCNTDTYEV-BENGVRKCKKCDGLCSK 342

RESULT 5
EGFR_HUMAN
ID EGFR_HUMAN STANDARD: PRT: 1210 AA.
AC P00533; P06268; Q14225;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84219729; Pubmed=6328312;
RA Ullrich A., Coussens L., Haylick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayer E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RA "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; Pubmed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN [3]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; Pubmed=6330563;
RA Xu Y., Ishii S., Clark A.D.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RA "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN [4]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; Pubmed=6093780;
RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RL receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [5]
RP SEQUENCE OF 1-29 FROM N.A.

```

```

RX MEDLINE=88217333; Pubmed=3329716;
RA Haley J., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396(1987).
RN [6]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; Pubmed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [7]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; Pubmed=2991899;
RA Ishii S., Xu Y.H., Straton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [8]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [9]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; Pubmed=6325948;
RA Mroczkowski B., Mosis G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA.";
RL Nature 309:270-273(1984).
RN [10]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; Pubmed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HER2/neu are located in their carboxyl-terminal tails.
RL Identification of a novel site in EGF receptor.";
RL J. Biol. Chem. 264:10667-10671(1989).
RN [11]
RP REVIEW.
RX MEDLINE=87297456; Pubmed=3039909;
RA Carpenter G.;
RT "Receptors for epidermal growth factor and other polypeptide
RT mitogens.";
RL Annu. Rev. Biochem. 56:881-914(1987).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- MISCELLANEOUS: BINDING OF EGF TO THE RECEPTOR LEADS TO
CC INTERNALIZATION OF THE EGF-RECEPTOR COMPLEX, INDUCTION OF THE
CC TYROSINE KINASE ACTIVITY, STIMULATION OF CELL DNA SYNTHESIS, AND
CC CELL PROLIFERATION.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00588; CAA25240.1; -
DR EMBL; X06370; CAA29668.1; -
DR EMBL; X00663; CAA25282.1; -
DR EMBL; M38425; AAA63171.1; -
DR EMBL; M11234; AAA52370.1; -

```

DR PIR: A00641; GQHUE.
 DR PIR: A00642; GQHUE2.
 DR PIR: A23062; A23062.
 DR HSP: P11362; IFGI.
 DR SWISS-2DPAGE: P00533; HUMAN.
 DR MIM: 131550; -.
 DR InterPro: IPR000494; EGFR.L.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; Fu; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane: Glycoprotein; Repeat; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 645
 FT TRANSMEM 646 668
 FT DOMAIN 669 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT REPEAT 1025 600
 FT DOMAIN 1025 1071
 FT DOMAIN 712 979
 FT NP_BIND 718 726
 FT BINDING 745 745
 FT ACT_SITE 837 837
 FT MOD_RES 678 678
 FT MOD_RES 1092 1092
 FT MOD_RES 1110 1110
 FT MOD_RES 1172 1172
 FT MOD_RES 1197 1197
 FT CARBOHYD 128 128
 FT CARBOHYD 175 175
 FT CARBOHYD 196 196
 FT CARBOHYD 332 332
 FT CARBOHYD 361 361
 FT CARBOHYD 413 413
 FT CARBOHYD 444 444
 FT CARBOHYD 528 528
 FT CARBOHYD 568 568
 FT CARBOHYD 603 603
 FT CARBOHYD 623 623
 FT CONFLICT 540 540
 SEQUENCE 1210 AA; 134277 MW; DBA2A50B4EFB6ED2 CRC64;

Query Match 34.7%; Score 793; DB 1; Length 1210;
 Best Local Similarity 45.3%; Pred. No. 2.2e-54;
 Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

QY 11 LLLALLPPCAA--STOVCGTDMKRLRPA SPETHDMLTHVGCGOYVGNLETPPTN 68
 DB 14 LLAALCPASRALDEKKVCCGTSKLTQLCTFEDHPLSLDPMFNKNCVVLGNLEITYVGRN 73
 QY 69 ASLSFLDIOEVGYVLIHNOVROYPLQRLRIVRGSTQLEFNVALAVLDNGDPLNNTPT 128
 DB 74 YLSLFLKTIQEVAGYVLIHNOVRYPLQRLRIVRGSTQLEFNVALAVLDNGDPLNNTPT 126
 QY 129 VYGSGFGIREFLIQSLVLIHNOVRYPLQRLRIVRGSTQLEFNVALAVLDNGDPLNNTPT 188
 DB 127 ---ANTGTKEIPMLRLOIHLGAVFSSNNPALCNVESIOMDIYSSDLSNMSPDQNH 183
 QY 189 SNACHPGCMKSGRCWSSSEDCSLTRTVACAGCA -CKEGLPTDCCHEGCAACTGP 247
 DB 184 LSCQKCDKSPGSCWGSAGEECQVLTITLCAQCSGRCGRSPSPDCHNCAACTGP 243
 QY 248 KHSDCIACLFHNSGICEHCPALVLYNDFESMPNPGRYTFGASCVTACPYNYLSTPD 307

DB 244 RESDCLVCRKFRDEATCTDTPCLMLYNTPTTYQMNVNPEGKTSFGATCVKCKPRHYVTD 303
 QY 308 VGSCTVLPCLHNOEVTAEDGTQRCCKSCPCAR 340
 DB 304 HGSVCVACGADSEYM-EEGVCVKCKCKGCPCK 335
 RESULT 6
 EGF_MOUSE
 ID EGF_MOUSE STANDARD; PRT; 1210 AA.
 AC Q01279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112).
 GN EGF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAIB/C; TISSUE=Liver;
 RX MEDLINE=93026370; PubMed=1408137;
 RA Avioli A., Skorecki K., Yayon A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 (bek/Kcfr) gene.";
 RL Oncogene 7:1957-1962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAIB/C, AND CD-1; TISSUE=Liver, and Decidua;
 RX MEDLINE=93126380; PubMed=7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 in mouse blastocysts during delayed implantation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAIB/C; TISSUE=Liver;
 RX Hildes M.L.;
 RT Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170986; PubMed=8125255;
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Barp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 receptor tyrosine kinase.";
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE=Brain;
 RA Medvinsky A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
 RT "Comparison of EGF receptor sequences as a guide to study the ligand
 binding site.";
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RA Eisinger D.P., Serrero G.;
 RT Submitted (Jun-1992) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- MISCELLANEOUS: BINDING OF EGF TO THE RECEPTOR LEADS TO
 CC INTERNALIZATION OF THE EGF-RECEPTOR COMPLEX, INDUCTION OF THE
 CC TYROSINE KINASE ACTIVITY, STIMULATION OF CELL DNA SYNTHESIS, AND

CC CELL PROLIFERATION.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X78987; CAA55587.1; -
 CC EMBL: U03425; AAA1789.1; -
 CC EMBL: X59698; CAA42219.1; -
 CC EMBL: L06864; AAA53029.1; -
 CC EMBL: Z12608; CAA78249.1; -
 CC HSSP: P11362; IFC1.
 CC MGD: MGI:95294; Egfr.
 CC InterPro: IPR000494; EGFR_L.
 CC InterPro: IPR000719; Euk_pkinase.
 CC InterPro: IPR001274; Furin-like.
 CC InterPro: IPR001245; Tyr_kin.
 CC Pfam: PF00757; Furin-like; 1.
 CC Pfam: PF00069; pkinase; 1.
 CC Pfam: PF01030; Recep_L_domain; 2.
 CC SMART: SM00261; Fv; 3.
 CC SMART: SM00261; Tyrc; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 CC Trasnembrane: Glycoprotein; Receptor; Signal; Transferase;
 CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 CC SIGNAL 1 24
 CC CHAIN 25 1210
 CC DOMAIN 25 647
 CC TRANSSEM 648 670
 CC DOMAIN 671 1210
 CC REPEAT 75 300
 CC REPEAT 390 600
 CC DOMAIN 1028 1071
 CC DOMAIN 714 981
 CC NP_BIND 720 728
 CC BINDING 747 747
 CC ACT_SITE 839 839
 CC MOD_RES 680 1092
 CC MOD_RES 1092 1092
 CC MOD_RES 1110 1110
 CC MOD_RES 1172 1172
 CC MOD_RES 1197 1197
 CC CARBOHYD 128 128
 CC CARBOHYD 175 175
 CC CARBOHYD 196 196
 CC CARBOHYD 352 352
 CC CARBOHYD 413 413
 CC CARBOHYD 444 444
 CC CARBOHYD 528 528
 CC CARBOHYD 568 568
 CC CARBOHYD 603 603
 CC CARBOHYD 623 623
 CC CONFLICT 19 19
 CC CONFLICT 539 539
 CC CONFLICT 991 991
 CC CONFLICT 1116 1117
 CC SEQUENCE 1210 AA; 134853 MW; 690E20D46DFD22F5 CRC64;

Db 14 LFTALCAAGALAEKRYKQCGTSNRLTGTEDHFLSLQRYNNCEVVLGNLEITTYQVRN 73
 Oy ASLSFLDIOIEVOGVYLANHNOVROVPLORIRYRGTOLEFDNYALAVLDGDDPLNNTTP 128
 Db 74 YDLSFKTIQEVAGVYVLAIVLMTVERIPENLOITRGALNNTYALATLSN----- 124
 Oy 129 VTGASPGELRELOLRLSTLEILKGVLJORNOLCYODTILMKDI-----FHNNOALATLI 184
 Db 125 -YGNRTGRLRELPRNRIQELLIGAVRSNNILCNMDPTIQWRDLYOVNFMNMDL--- 180
 Oy 185 DTNRSRACHPCSPMKSGRCGSESEDCSLTRIVCAGCA-RCKGLPTDCCHEQCAAG 243
 Db 181 -QSHPSKCPKCDPSCPNKSCWGGEENCQKLTIKICAQOCSHRRCGRSPSCCHNOCAAG 239
 Oy 244 CTGPKHSKCLACLFHNSGICELHCPALVYVNTDFEESMPREGRYFGASCYACACRYN 303
 Db 240 CTGPRESKCLVQKFRQDEATKDTCPPLMLNPTTYQMDVNPBGRKYSFGATCAVKKCPRY 299
 Oy 304 LSTVGSCTLVCPHNDVETADGTORCEKSKPCAR 340
 Db 300 VVDHGSVCVRACGPDYIEV-BEDGIRKCKKCDGFCRK 335
 RESULT 7
 ERB4_HUMAN STANDARD; PRT; 1308 AA.
 ID ERB4_HUMAN
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ERB4-4 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
 DE (P180ERB4) (TYROSINE KINASE-TYPE CELL SURFACE RECEPTOR HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE-Breast carcinoma;
 RX MEDLINE=93189574; PubMed=8383326;
 RA Plozman G.D., Culouscou J.-M., Whitley G.S., Green J.M., Carlton G.W.,
 RA Foy L., Neubauer M.G., Shoyab M.;
 RT "Ligand-specific activation of HER4/p180erb4, a fourth member of the
 RL epidermal growth factor receptor family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE-Fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Cortas G., Paul S., Choi C.J., Rio C., Plozman G.D.,
 RA Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/erb4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester.";
 RL J. Biol. Chem. 272:26761-26768(1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,

[illegible]

Oy	125	NTTPTVTGASGCGRELOLRSLEITLFGVLLKGVLOLPCYOTLIILMKDIFHKNNOALATLTI	184
Db	128	NE-----GIGLEIAGKMLFTILLNGVYVQDNKFLCYADTIHQDILVRPWPNSNLTLY	178
Oy	185	DTNRSRACHPCSPWCKSGRMSGSESSDQSLTRTYCAGGC-ARCKGPLPTDCHBQCAAG	243
Db	179	STNSSSGCGRCHKSCSTG-RWCGPTENHCOTLTRFYCAEQCDQRCRGYPVSCCHRECAAG	237
Oy	244	CTGKHSIDCLACHFNHNSGICELHCPALTYNTDHFESMPNPEGXYTGASCACVACPYNY	303
Db	238	CSGKRDIDCEACAKMNPDSGACVYQCQTFVYVPTTFQLEHNHNAVITYTGACVAKKCPHNF	297
Oy	304	LSTVGSCTLVCEPLHNOEVAEDQTRCEKSCPPAR-----GTHSLT	346
Db	298	V-VDSSCVACRCPSSKMEV-EEENGIMCKRPTCIDICKACDGIIGTOSLM	343
RESULT	8		
ERB3_HUMAN			
ID	ERB3_HUMAN	STANDARD:	PRT: 1342 AA.
AC	P21860:		
DT	01-MAY-1991 (Rel. 18, Created)		
DT	01-MAY-1991 (Rel. 18, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	ERBB-3 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)		
DE	(TYROSINE KINASE-TYPE CELL SURFACE RECEPTOR HER3).		
OS	ERBB3 OR HER3.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxId:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90883234; PubMed=2687875;		
RA	Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;		
RA	Plozman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,		
RA	Todaro G.J., Shoyab M.;		
RT	"Molecular cloning and expression of an additional epidermal growth		
RT	factor receptor-related gene.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).		
RN	[3]		
RP	SEQUENCE FROM N.A. (SHORT FORM).		
RP	TISSUE=Placenta;		
RX	MEDLINE=93282822; PubMed=7685162;		
RA	Katch M., Yazaki Y., Sugimura T., Terada M.;		
RT	"c-erbB3 gene encodes secreted as well as transmembrane receptor		
RT	tyrosine kinase.";		
RL	Biochem. Biophys. Res. Commun. 192:1189-1197(1993).		
CC	-1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTFAK.		
CC	-1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE -> ADP +		
CC	PROTEIN TYROSINE PHOSPHATE.		
CC	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS		
CC	(POTENTIAL).		
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND		
CC	SECRETED (SHORT FORM).		
CC	-1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A		
CC	SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.		
CC	-1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.		
CC	-1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE		
CC	SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.		
CC	-1- PTM: LIPID-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES		
CC	AND PROMOTES ITS ASSOCIATION WITH THE B5 SUBUNIT OF		
CC	PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).		
CC	-1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.		
CC	-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.		

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M29366; AAA35790.1; -
DR EMBL: M34309; AAA35979.1; -
DR EMBL: S61953; AAB26935.1; -
DR PIR: A36223; A36223.
DR HSSP: P1362; 1FC1.
DR MIM: 190151; -
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-kin.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KM Alternative splicing.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1342 ERBB-3 RECEPTOR PROTEIN-TYROSINE KINASE.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 664 POTENTIAL.
FT DOMAIN 665 1342 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 709 966 PROTEIN KINASE.
FT NP_BIND 715 723 ATP (BY SIMILARITY).
FT BINDING 742 742 ATP (BY SIMILARITY).
FT ACT_SITE 834 834 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDKLCHEMDTIDMDRIYDRDAEIVKDNGR
SC -> GQFPMVPSGLTFQPADMDRIYDRDAEIVKDNGR
VPTLTAIV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

Query Match 32.2%; Score 735.5; DB 1; Length 1342;
Best Local Similarity 44.0%; Pred. No. 7.7e-50;
Matches 150; Conservative 49; Mismatches 121; Indels 21; Gaps 9;

```

```

OY 188 RSACHPSPCKSGSRGCESEDCOSLTRYACAGG--ARCKGLPTDCCHROCAAGCTG 246
DB 179 NGRSCPCHVEYCKG--RCWPGSEDCQTLTKITCAQCNGHGGPNPNOCCDCEAGGSG 237
OY 247 PKHSQCLACILHNHSGICELHCPALVTYNTPFEESMPNDEGRYTGASCVTPACPNYST 306
DB 238 PDDTCCFACRHNHNDGCAVCPRCPPLVYNNKLTFOLEPRLPHIKYQYGVCAVASCPRNFV-V 296
OY 307 DVGSCTLVCPHNDQEVTAEDGTORCEKCSKPCAR--GTHS 344
DB 297 DQTSVCRACPDPMKEVD--KNGLKMCPECGGLCPKACEGTGS 336

RESULT 9
ERB3_RAT STANDARD: PRT; 1339 AA.
AC 062799; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ERBB-3 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112).
GN ERBB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sterke S.L., Roland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein.";
RL Gene 165:279-284(1995).
RN [2]
RP REVISION TO 85.
RA Hellyer N.J.;
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neurogulin and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U29339; AAC28498.1; -
DR EMBL: U52530; AAC53050.1; -
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.

```

DR Interpro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep.L.domain; 2.
 DR PRINTS: PRO0109; TYRKINASE.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane, Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 333 333
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT CONFLICT 1028 1028
 SO SEQUENCE 1339 AA; 147619 MW; F120281D432423D8 CRC64;

Query Match 30.5%; Score 698; DB 1; Length 1339;
 Best Local Similarity 42.1%; Pred. No. 6, 6e-47;
 Matches 147; Conservative 46; Mismatches 130; Indels 26; Gaps 10;

QY 3 LAAALCWGLLALLPPGAA---STVCCTGTDKMLRLPASPETHLMDLRLLXQCCVVOGN 59
 DB 7 LQVLC---FLSLIARGSEMGNSQAQVCPGLTNGSLVTDGADNOYQTLVLYKCEVWGN 62
 QY 60 LETVYLPNALSFLDIOEVQGVYLIANOVQVPLQRLRIYRGQLFEDNYALVLN 119
 DB 63 LELVLGNHADSFLQWIEVGYVYVAMNEFVLPRLRYRGQVYDGAFAIVM-- 120
 QY 120 GDLNNTTPTVGTASPGGLRELRLSITELIKGVLIQRPQCLCYDTILMKDIFHKNNQ 179
 DB 121 ---LNYNT---NSSHALRQLKFTQLEILSGVYTEKDKLCHMDTIDMRDIVRVR-- 170
 QY 180 ALTLIDTNSRACHPCSPCKSGRCGSESESDOSLTLTVVAGGC-ARKGPIPTCCHE 238
 DB 171 GAETIVKKNGANCPCHVECKG-RCWGPGRDDQILTKTICAPQCNGRGFPNQCCHD 229
 QY 239 OCAAGCTGPKHSKDLACLHFNHSGICELCPALVYNTPTFESMPNDECRYFGASCYTA 298
 DB 230 ECAGGCGSGQDPDCCFACRFNFDSGACVPRCPRELYVNNKLTFLQEPRLPHRIKYGGCVAS 289
 QY 299 CPYNYISTDVGSCITLVCPILHNOEVTAEDETQRCCKSKPCAR---GTHS 344
 DB 290 CPNHFV-VDQTEFCVRACPPDKMEVD-KHGLKMCPCGGLCPKRCGEGTS 336
 RESULT 10
 XMRK_XIPMA STANDARD; PRT; 1167 AA.
 AC P13388;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MELANOMA RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112).
 GN XMRK OR TU.

OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorphi; Atherinomorphi;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Witbrodt J., Adam D., Maltzschek B., Meunier W., Raulf F.,
 RA Telling A., Robertson S.M., Scharl M.,
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing tu locus in Xiphophorus.";
 RL Nature 341:415-421(1989).
 RP REVISION TO 515.
 RA Scharl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X16891; CAA34770.2; -
 DR PIR: S06142; S06142.
 DR HSSP: P1362; IEG1.
 DR Interpro: IPR000494; EGFR.L.
 DR Interpro: IPR000719; Euk_pkinase.
 DR Interpro: IPR002174; Furin-like.
 DR Interpro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep.L.domain; 2.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane, Glycoprotein; Receptor; Signal; Transferrase;
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167
 FT DOMAIN 26 642
 FT TRANSMEM 643 665
 FT DOMAIN 666 1167
 FT DOMAIN 710 927
 FT NP_BIND 716 724
 FT BINDING 743 743
 FT ACT_SITE 835 835
 FT CARBOHYD 114 114
 FT CARBOHYD 144 144
 FT CARBOHYD 201 201
 FT CARBOHYD 356 356
 FT CARBOHYD 365 365
 FT CARBOHYD 398 398
 FT CARBOHYD 417 417
 FT CARBOHYD 501 501
 FT CARBOHYD 576 576
 FT CARBOHYD 621 621
 SO SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 29.8%; Score 661.5; DB 1; Length 1167;
Best Local Similarity 42.4%; Pred. No. 1.1e-45;
Matches 145; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

```
OY 4 AALCAGMGLLALLPPGAAST---OVCGTQDMLKLPASPFHMLRLHYGGCOVGN 59
DB 8 AALLO--LLLVLSIRCCSTDPDRKVCOSTSNOMT---LDNHYTKMKMMSGCNVLEN 62
OY 60 LELVLPNTNASLFLQDIOEVGYVLLAHNOYQVPLQRLRIVRGSTOLFEDNYALAVLDN 119
DB 63 LEIVTQENODLSFIQIOEVGYVLLAMNEVSTIPVNLRLIRGONLYEGFILLVMSN 122
OY 120 GDPPLNNTPIYVIGASGGLRELDLSLTCLKGVLIQRNPOLCYODTIMKIFIKNNOL 179
DB 123 YQK-NPSS--DYVYQGLKQLQSLMTLETSGVAVSHNPPLCANNETIMMDIVKTSNP 179
OY 180 ATLTLDTNRSRACHCSPMKSGRCMGESSSEDCOSLTFRYCAGGC-ARCCKGLPTDCHE 238
DB 180 TMLNLPHPAFEROCCKCHGCVNGSCWMAQPGHCQFTKLCLAEQCNRRKGRKPIDCNE 239
OY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYTYNTDFESMNPREGRYFGASCYA 298
DB 240 HCAGGCTGPRATDCLACDFNDGCKDPCPPKLYDLYSHQVYVNDPNKTYFFGAACVKE 299
OY 299 CPYNTLSTDVSGCTLVCPPLHNOVTAEDGTQCEKSCRPAC 340
DB 300 CPSTNYVTE-GACVRSCSAGMLEVD-ENGKRCCKPCDGVCPK 339

RESULT 11
BGRF_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112) (EGFR)
DE (GRIKIN RECEPTOR) (TORPEDO PROTEIN) (DROSOPHILA RELATIVE OF ERBB).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RP MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
[2]
RN REVISIONS.
RP Clifford R., Schubach T.;
RL submitted (Mar-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Lynen E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains."
RL Cell 40:599-607(1985).
[4]
RN SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=OREGON-R; TISSUE=Embryo;
RA Medline=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts."
RL Cell 46:1091-1101(1986).
[5]
RN SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II), AND CHARACTERIZATION.
RP
```

RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RL specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).

RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=BKBERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Fandel M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abbil J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Baller R.M., Basu A.P., Benham A., Benham B.P., Bhandari D., Bolshakov S.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Bothan M.R., Boutek J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavellier S., Dahlke C., Davenport L.B., Davies P.,
RA de Pallos B., Delcher A., Deng Z., Days A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston F.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liu X.P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liao X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,
RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shee B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svyrkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).

RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=DAEKWANYEONG;
RX MEDLINE=85137938; PubMed=2983232;
RA Masworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal
RL growth factor receptor.";
RL Nature 314:178-180(1985).

RN [8]
RP PARTIAL SEQUENCE FROM N.A., AND MUTATION ANALYSIS.
RC MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among DER/flb alleles: implications for
RL the mechanism of signal transduction by receptor-tyrosine kinases";
RL Genetics 129:191-201(1991).

RN [9]
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the Drosophila
RL EGF receptor.";
RL Cell 89:13-16(1997).

CC -I- FUNCTION: BLINDS NO FOUR LIGANDS: SPITZ, GUREN, VEIN AND ARGOS,
WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF
MAPK PATHWAY. INVOLVED IN A WRAPD OF DEVELOPMENTAL DECISIONS.
CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE

CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEOSA
 CC AND VENTRAL NEUROECODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS, IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, HIGHEST
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL AF052754 AAC08536.1 -
 DR EMBL AF052753 AAC08536.1 JOINED.
 DR EMBL AF052754 AAC08535.1 -
 DR EMBL AF052752 AAC08535.1 JOINED.
 DR EMBL K03054 AA551462.1 -
 DR EMBL K03417 AA551460.1 -
 DR EMBL K03416 AA550965.1 -
 DR EMBL K03418 AA551461.1 -
 DR EMBL AF109077 AAD26134.1 -
 DR EMBL AF109078 AAD26132.1 -
 DR EMBL AF109082 AAD26132.1 JOINED.
 DR EMBL AF109078 AAD26133.1 -
 DR EMBL AF109078 AAD26133.1 JOINED.
 DR EMBL AF109079 AAD26130.1 -
 DR EMBL AF109081 AAD26130.1 JOINED.
 DR EMBL AF109079 AAD26131.1 -
 DR EMBL AF109083 AAD26131.1 JOINED.
 DR EMBL AF109080 AAD26135.1 -
 DR EMBL AE003454 AAF46732.1 -
 DR EMBL X02293 CA26157.1 -
 DR EMBL X78920 CA555523.1 -
 DR EMBL X78919 CA555521.1 -
 DR EMBL X78919 CA555522.1 -
 DR PIR A00640 GQFEF.
 DR HSPF P11362 IFEI.
 DR FLYBASE FBgn0003731. Egr.
 DR InterPro: IPR000494; EGRF.L.
 DR InterPro: IPR000719; EGRF.L.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam PF00757; Furin-like; 1.
 DR Pfam PF00069; pkinase; 1.
 DR Pfam PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 7.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30
 FT CHAIN 31 1426
 FT POTENTIAL.
 FT EPIDERMAL GROWTH FACTOR RECEPTOR.

FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 869 889 POTENTIAL.
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 938 1198 PROTEIN KINASE.
 FT NP_BIND 944 952 ATP (BY SIMILARITY).
 FT BINDING 971 971 ATP (BY SIMILARITY).
 FT ACT_SITE 1063 1063 BY SIMILARITY.
 FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 FT MOD_RES 1309 1309 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 128 128 N-LINKED GLCNAC. . . (POTENTIAL).
 FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
 Query Match 25.1%; Score 574.5; DB 1; Length 1426;
 Best Local Similarity 36.3%; Pred. No. 3.2e-37;
 Matches 118; Conservative 46; Mismatches 130; Indels 31; Gaps 7;
 QY 24 QVSTGDMKRLPLASPEYHLDMLRLHYGCGVAGNLELYLPT-NASLSFLQDIQEVQG 82
 DB 100 KICIGTSRLSPVSKNKHRYNRLRDRTNCTYVDGNLKLWLPNENLDLSFDNIREYVG 159
 QY 83 YVLIAHNVQVQPLQRLRIVRGTOLE---EDNYALAVLDNGDPLNMTTPYTGASPGGL 137
 DB 160 YLLISHVDKRVKVPKRLQIRGRTLSLSVEEKTALFY-----TYSKM 203
 QY 138 RELQLRSLTEILKGVLLQIQRNPOLCYODTILKKDIFKKNQDLATLIDTNRSRACHPCSP 197
 DB 204 YLLEIPDLRDVLNGVGFNNVNLCHMRTIQSELSVNGTDAYVYVDFAPARECPKCHE 263
 QY 198 MCKSGRCWGESSEDCQSLRTIYACAGCA--RCKGPLPDDCCHGQACAGTGKHSQCLAC 255
 DB 264 SCTHG-CWGEGRPKKQKSKLCSPOCAGRCYGRPCCHLFCAGCTGPTOKDCLAC 322
 QY 256 LHPNHSGLCELPALVYNTDTPESMPNPEGRTYFGASCYACPYNYLSTDVGSCTLYVC 315
 DB 323 KNFDEAVSKCECPKPKRYNPTTYVLETNPEKIVYGVATCYACEP-GHLRDNGACVNSC 381
 QY 316 PLHNOEVTAEIDSTORCEKSKPCAR 340
 DB 382 PQDKMDKGE-----CVPNGPCPK 401
 RESULT 12
 ID LT23_CAEEL STANDARD; PRT; 1323 AA.
 AC P23448;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE LET-23 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112).
 GN LET-23 OR KIN-7 OR ZK1067.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Rhabditinae; Caenorhabditis.
 OC NCBL_TaxID=6239;
 OX 11
 RN RN
 RP SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RX MEDLINE=91080919; PubMed=1979659;
 RA Aroian R.V., Koga M., Mendel J.E., Ohshima Y., Sternberg P.W.;
 RT "The let-23 gene necessary for Caenorhabditis elegans vulva
 RT induction encodes a tyrosine kinase of the EGF receptor subfamily.";
 RL Nature 348:693-699(1990).
 RN 12
 RP SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RX MEDLINE=96177760; PubMed=8604137;
 RA Sakai T., Koga M., Ohshima Y.;
 RT "Genomic structure and 5' regulatory regions of the let-23 gene in
 RT the nematode C. elegans.";
 RL J. Mol. Biol. 256:548-555(1996).
 RN 13
 RP SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;

RA Thomas K.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP MUTANTS.
 RX MEDLINE-94147981; PubMed-8313860;
 RA Aroian R.V., Les G.M., Sternberg P.W.;
 RT "Mutations in the *Caenorhabditis elegans* let-23 EGFR-like gene define
 RT elements important for cell-type specificity and function."
 RL EMBL J. 13:360-366(1994).
 CC -!- FUNCTION: TYROSINE KINASE RECEPTOR REQUIRED FOR THE INDUCTION OF
 CC C.ELEGANS VULVA. POSSIBLE RECEPTOR FOR THE INDUCTIVE SIGNAL
 CC REQUIRED FOR VULVAL DEVELOPMENT. ACTIVATED BY LIN-3 AND ACTS BY
 CC WAY OF LET-60 RAS. THE LIN-3/LET-23 PAIR IS A SIMPLIFIED VERSION
 CC OF THE MAMMALIAN NEUREGULIN-ERBB NETWORK.
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X57767; CAA00919.1; ALT_SEQ.
 DR EMBL: D63426; BAA09729.1; ALT_INIT.
 DR EMBL: 270038; CAA93882.1; -.
 DR PIR: S13422; S13422.
 DR HSSP: P1362; 1FG1.
 DR WormPep: ZK1067.1; CE03840.
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 6.
 DR SMART: SM00219; TyrcK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Transmembrane; Glycoprotein; Receptor; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 1323
 FT DOMAIN 23 818
 FT TRANSMEM 819 841
 FT DOMAIN 842 1323
 FT DOMAIN 885 1152
 FT NP_BIND 891 899
 FT BINDING 919 919
 FT ACT_SITE 1010 1010
 FT CAROHRD 91 91
 FT CAROHRD 169 169
 FT CAROHRD 255 255
 FT CAROHRD 376 376
 FT CAROHRD 561 561
 FT CAROHRD 655 655
 FT CAROHRD 746 746
 FT CAROHRD 776 776
 FT VARIANT 469 469
 FT VARIANT 700 700
 FT VARIANT 753 753
 FT VARIANT 1065 1065
 FT VARIANT 1074 1074
 FT VARIANT 1323 AA; 150510 MW; 6B0307E53EEFA99 CRC64;
 SQ SEQUENCE

Query Match 18.3%; Score 419; DB 1; Length 1323;
 Best Local Similarity 29.0%; Pred. No. 4.3e-25;
 Matches 102; Conservatives 59; Mismatches 131; Indels 60; Gaps 14;
 QY 25 VCTGTDMLRLPASPEHIDMLRHLYGCGVVOGNELETYLPN----- 68
 DB 39 LSGGTNGISRGVGTGNI-LEDELTWYRGCRVYNGNEIWEANEIKKWRSTNSTVDK 97
 QY 69 -----ASISPLDIDIOVGQYVLIHNOYRQVPLQRLRYRGTQTFEDYALAVDNGCP 122
 DB 98 NEDSPKLSINFDNLEIGSLIYRANIQKISFPLRYVYDEVEFDN-ALYIHNDR- 155
 QY 123 LNNITFVTGASPGGURELQRLTEILKGVLIQRPOCY-ODTFLMKDIFHKNQQL 181
 DB 156 -----VHEVYMRRLRYRNSVYTIQDNPKCYIGDKIDMKELLYDPD--VQ 199
 QY 182 TLIDNRSRACH-----PCSPMKSGSRWESSEDCSLTRFYCAGGCARC---KGPL 231
 DB 200 KVETNSHQHCYONGKSMKCHESC-NDKCGSGDNDGQRYRVRYCPSKSCQCFVSNST 258
 QY 232 PDDCHGCGACGCTGKSHSDCLAFHNSGICELHCPALVYVNDTFESMPNPGRTYF 291
 DB 259 SYECDSDACIGCTGHPKNCIACSKYELDGIETCPSPKIFNHTGRVLVFNPDGRYON 318
 QY 292 GASCVTACPYNTL-STDVGSCTLYC-PLHNOEYTAEDGQROEKC-SKPCAR 340
 DB 319 GHNCHVKECPPELLIENDV--CYRHCSGDGHYDITKD--VRECEKNCSSCPK 366
 RESULT 13
 ID ILPR_BRALA STANDARD; PRT; 1363 AA.
 AC 002466;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 DI INSULIN-LIKE PEPTIDE RECEPTOR PRECURSOR (EC 2.7.1.112) (ILP RECEPTOR).
 OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxId-7740;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96408719; PubMed-8813726;
 RA Pashmiforoush M., Chan S.J., Steiner D.F.;
 RT "Structure and expression of the insulin-like peptide receptor from
 RT amphioxus."
 RL Mol. Endocrinol. 10:857-866(1996).
 CC -!- FUNCTION: THIS RECEPTOR BINDS TO THE INSULIN RELATED PEPTIDE AND
 CC HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -!- SUBUNIT: PROBABLE TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY
 CC DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF
 CC THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAIN CARRY THE KINASE
 CC DOMAIN (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
 CC PROTEIN KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: S83394; AAB50848.1; -.
 DR HSSP: P06213; IIRK.
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001777; FN_III.

DR Interpro: IPR002174; Furin-like.
 DR Interpro: IPR002011; Receptor_tyr_kin_II.
 DR Pfam: PF00041; fn3; 3.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinese; 1.
 DR Pfam: PF01030; Recep_L_domain; 1.
 DR PRINTS: PR00014; FNTYPEI1.
 DR PRINTS: PR00109; TYRKINASE.
 DR SMART: SM00060; FN3; 3.
 DR SMART: SM00261; FU; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00239; RECEPTOR_TYR_KIN_II; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;
 KW Glycoprotein; ATP-binding; Phosphorylation; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 716
 FT PROPEP 717 720
 FT CHAIN 721 1363
 FT DOMAIN 721 928
 FT TRANSMEM 929 949
 FT DOMAIN 950 1363
 FT DOMAIN 994 ?
 FT NP_BIND 1000 1008
 FT BINDING 1028 1028
 FT ACT_SITE 1148 1148
 FT MOD_RES 1174 1174
 FT CARBOHYD 51 97
 FT CARBOHYD 97 97
 FT CARBOHYD 137 137
 FT CARBOHYD 278 278
 FT CARBOHYD 483 483
 FT CARBOHYD 599 599
 FT CARBOHYD 617 617
 FT CARBOHYD 665 665
 FT CARBOHYD 666 666
 FT CARBOHYD 711 711
 FT CARBOHYD 732 732
 FT CARBOHYD 736 736
 FT CARBOHYD 743 743
 FT CARBOHYD 816 816
 FT CARBOHYD 885 885
 FT CARBOHYD 888 888
 FT CARBOHYD 898 898
 FT CARBOHYD 1363 1363
 FT SEQUENCE 1363 1363
 FT AA: 154104 MW: 23812084 EABLE1D65 CRC64;

Query Match 15.0%; Score 342.5; DB 1; Length 1363;
 Best Local Similarity 28.3%; Pred. No. 4.3e-19;
 Matches 106; Conservative 40; Mismatches 124; Indels 105; Gaps 20;

QY 9 WGLI-----LALIPGAASTQVCTGDMKRLPASPEHILDMRLHYQCGVQGMILELT 63
 DB 12 MAALTIVTIGLIVPSNGEYICDSMDIRN-----VSNRLQI-ENCTVIEGYLQI- 61
 QY 64 YLPTNLSLTDIOIVQGVLAHNOVYV-----LQRLR-----IYRGV 106
 DB 62 -----LLIDFAEODSGSLAPNVLVEITYFLLYRGRGLTNLSELFPNLAVIRGN 112
 QY 107 LFDENYALAVINDGDEPLNNTPTVPGASPGGLRELQRLSEILKGVLIQNRNOLCYODT 166
 DB 113 LF-FNVALVFEKMD-----MOKIGLXSLQNTIRGSRVRIEKNKNLCYLDTP 156
 QY 167 ILMKDI---HKNNQALALTLIDNRSRAC-HPCSPMK-----GSRCEWSESSDCOSLT 216
 DB 157 IDMSFIAESGYNN-----FIVDNREEBECVNFPCGRKRIKHPVLIQDLCWAE--EHCQKVC 210
 QY 217 RIVYAGGACARCKPPLPDCCHCECAAGCTGPKKSDCLACHFNHSGICELHCAALTYNT 276

DB 211 PESICGNCR---DGISCCHENCIGGCDPTERDCVACKFVANGECILQCPPTIYK 266
 QY 277 D-----TFESMPNPEGRY--TFGASCVTACPYNYLISTDVG---CTLYCPLHNOEYTAEDG 327
 DB 267 DRACITEECPCPTNTSVKMLHHRKICRCPSPGY--TTDLNPNRLCT----- 310
 QY 328 TORCE-KCSKPCARG 341
 DB 311 --EEGOCPCSKCKG 323
 RESULT 14
 INSR_DROME STANDARD; PRT; 2146 AA.
 AC P09208; Q24089; Q24023;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE INSULIN-LIKE RECEPTOR PRECURSOR (EC 2.7.1.112).
 GN INR OR INR-A OR DIR-A.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX MBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95354655; PubMed=7628438;
 RA Fernandez R., Tabarini D., Azpiroz N., Frasch M., Schllessinger J.;
 RT "The Drosophila insulin receptor homolog: a gene essential for
 RT embryonic development encodes two receptor isoforms with different
 RT signaling potential.";
 RL EMBO J. 14:3373-3384(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95181404; PubMed=7876183;
 RA Ruan Y., Chen C., Cao Y., Garofalo R.S.;
 RT "The Drosophila insulin receptor contains a novel carboxyl-terminal
 RT extension likely to play an important role in signal transduction.";
 RL J. Biol. Chem. 270:4236-4243(1995).
 RN [3]
 RP SEQUENCE OF 652-1749 FROM N.A.
 RC STRAIN=OREGON-R; TISSUE=Embryo;
 RX MEDLINE=87100165; PubMed=3099787;
 RA Nishida Y., Hata M., Nishizuka Y., Rutter W.J., Ebina Y.;
 RT "Cloning of a Drosophila cDNA encoding a polypeptide similar to the
 RT human insulin receptor precursor.";
 RL Biochem. Biophys. Res. Commun. 141:474-481(1986).
 RN [4]
 RP SEQUENCE OF 1297-1595 FROM N.A.
 RX MEDLINE=86259667; PubMed=3014506;
 RA Petruzzelli L., Herrera R., Arenas-Garcia R., Fernandez R.,
 RA Birbaum M.J., Rosen O.M.;
 RT "Isolation of a Drosophila genomic sequence homologous to the kinase
 RT domain of the human insulin receptor and detection of the
 RT phosphorylated Drosophila receptor with an anti-peptide antibody.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4710-4714(1986).
 CC -1- FUNCTION: THIS RECEPTOR PROBABLY BINDS AN INSULIN RELATED PROTEIN
 CC AND HAS A TYROSINE-PROTEIN KINASE ACTIVITY (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: Tetramer of 2 alpha and 2 beta chains linked by disulfide
 CC bonds. The alpha chains contribute to the formation of the ligand-
 CC binding domain, while the beta chain carry the kinase domain.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
 CC PROTEIN KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

[1]
RN SEQUENCE FROM N.A.
RA Steele R.E., Mai N.H., Lieu P., Shenk M.A.:
RL Submitted (Apr-1991) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS RECEPTOR PROBABLY BINDS AN INSULIN RELATED PROTEIN
CC AND HAS A TYROSINE-PROTEIN KINASE ACTIVITY (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN DIVIDING EPITHELIAL CELLS.
CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M64612; AAA68205.1; -
DR HSSP: P06213; IIRK.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_Dkinase.
DR InterPro: IPR001777; FN_III.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002011; Reptor_tyr_kin_II.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00041; In3; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Kinase; 1.
DR Pfam: PF01030; Recep_L_domain; 1.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00261; FU; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
KW Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;
KW Glycoprotein; ATP-binding; Phosphorylation; Signal.
FT SIGNAL 1 24
FT CHAIN 25 1477
FT DOMAIN 25 980 PUTATIVE INSULIN-LIKE PEPTIDE RECEPTOR.
FT TRANSMEM 981 1001 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1002 1477 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1044 1058
FT NP_BIND 1050 1058 ATP (BY SIMILARITY).
FT BINDING 1077 1077 ATP (BY SIMILARITY).
FT ACT_SITE 1175 1175
FT MOD_RES 1201 1201 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 55 55 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 255 255 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 325 325 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 549 549 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 644 644 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 732 732 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 791 791 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 895 895 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 957 957 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 1477 AA; 168276 MW; 74ACDBA7C6DEID41 CRC64;

QY 51 QGCVVOGNT---ELTYLPINASLSFLQDIOEVQGYVLIAHNOYRQVPL-QRLRIYRGTO 106
Db 54 QNCTCMHGNIIVKSKSTKYDEENFKPYEPKRLREITGYLLISCTLKFPPLPGTLVINGGD 113
QY 107 LFEONYALAVLDNDGPLNNTPTVPGASPGGRLQLSLPFIKGLVLIQRNPQLCYODT 166
Db 114 LIL-NYALVIYN-----EIKVEYFPSLRAILNGVHIGRNRLCYVNT 156
QY 167 ILMRDI---FHKNNQALATLIDTNRRA-----CHPC-----SPMCKSGRCWG---- 206
Db 157 IRMSIILKIDHQGTQGYLYL-ESNKLNDLGLCKHGCHPAPAGHDGP--KAQYCMGDPK 213
QY 207 --ESEDQSLTRIVYAGGACARCKGRLPTDCCHQCAAGCTGPRHSD-CLACLHF--NHS 261
Db 214 KQNKKAQCQRFCNTQC--GREGCLDGSDDHLCNHECLGGCSAINSTYNTCHACRYRIKST 271
QY 262 GICELHCPALVTYNDTF---ESMP-----NPGRYTFGACVYACPYNTLSTDVSGCT 312
Db 272 GQCVSKCP-RQYLVDRFLCQESCPYMSINSTYHNYLMQGEVTKCPYNTIS----- 323
QY 313 LVCPILNQEVTAEDGTORCEKC 334
Db 324 -----NNQ-----TKRCEKC 333

Search completed: April 11, 2002, 09:34:23
Job time: 310 sec

Query Match 12.7%; Score 291; DB 1; Length 1477;
Best Local Similarity 29.2%; Pred. No. 5,1e-15;
Matches 94; Conservative 34; Mismatches 114; Indels 80; Gaps 18;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:35:06 ; Search time 40.45 Seconds
(without alignments)
1515.157 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287
Sequence: 1 MELALACRWGLLALLPAGA.....VGRGPPDAHVAVNLSRYEG 419

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2275	99.5	419	09UK79	09UK79 homo sapien
2	1716.5	75.1	1259	6 018735	018735 canis famill
3	794	34.7	405	13 090836	090836 gallus gall
4	793	34.7	405	4 092795	092795 homo sapien
5	793	34.7	628	4 09H2C9	09H2C9 homo sapien
6	793	34.7	657	4 014226	014226 homo sapien
7	793	34.7	705	4 09G2X1	09G2X1 homo sapien
8	789	34.5	643	11 09ERV6	09ERV6 mus musculu
9	789	34.5	655	11 09WVF5	09WVF5 mus musculu
10	789	34.5	1210	11 09EP98	09EP98 mus musculu
11	781.5	34.2	1209	11 09QX70	09QX70 rattus norv
12	779.5	34.1	1308	11 09QZ27	09QZ27 rattus norv
13	773	33.8	473	11 09ESE0	09ESE0 rattus norv
14	734	32.1	331	4 09BUD7	09BUD7 homo sapien
15	723	31.6	149	6 09BG66	09BG66 oryctolagus
16	693	30.3	1165	13 09YH40	09YH40 xiphophorus
17	662.5	29.0	1328	13 P79754	P79754 tuagu rubrip
18	647	28.3	599	13 09PSH2	09PSH2 gallus gall
19	571.5	25.0	1433	5 09BIR9	09BIR9 anopheles g

20	463.5	20.3	1137	13 09W6F6	09W6F6 gallus gall
21	404.5	17.7	150	6 09BG64	09BG64 oryctolagus
22	402.5	17.6	1368	5 023821	023821 caenorhabdi
23	395	17.3	151	6 09BG65	09BG65 oryctolagus
24	365.5	15.9	366	5 026569	026569 schistosoma
25	363.5	15.9	1717	5 026566	026566 schistosoma
26	331	14.5	334	5 026567	026567 schistosoma
27	331	14.5	342	5 026568	026568 schistosoma
28	300	13.1	1472	5 09U5A8	09U5A8 bombyx mori
29	298.5	13.1	1671	5 09NUT5	09NUT5 biotomalari
30	298	13.0	1358	13 073798	073798 xenopus lae
31	298	13.0	1418	13 093457	093457 scophthalmu
32	294	12.9	2144	5 09VD94	09VD94 drosophila
33	290	12.7	1300	11 09QVW4	09QVW4 mus musculu
34	284	12.4	1362	13 09PVZ4	09PVZ4 xenopus lae
35	273.5	12.0	89	11 088459	088459 mus musculu
36	272	11.9	469	11 063721	063721 rattus norv
37	270	11.8	410	11 063720	063720 rattus norv
38	252.5	11.0	1371	11 09QVW4	09QVW4 mus musculu
39	232	10.1	1245	13 09YGH8	09YGH8 scophthalmu
40	230	10.1	946	5 09VJ04	09VJ04 drosophila
41	210	9.2	868	5 03VER2	03VER2 drosophila
42	203	8.9	1846	5 016131	016131 caenorhabdi
43	191	8.4	82	6 09NOK4	09NOK4 sus scrofa
44	189	8.3	131	5 09BH16	09BH16 anopheles g
45	181.5	7.9	91	11 088458	088458 mus musculu

ALIGNMENTS

RESULT 1

ID 09UK79 PRELIMINARY; PRT: 419 AA.

AC 09UK79;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE HERSTATIN.

GN HER-2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxId=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99415951; PubMed=10485918;

RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;

RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted

RT autolnhibitor.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).

RN [2]

RP SEQUENCE FROM N.A.

RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF177761; AAD56009.2; -

DR InterPro: IPR000494; EGFRL.

DR InterPro: IPR002174; Furin-like.

DR Pfam: PF00757; Furin-like; 1.

DR Pfam: PF01030; Recep_L-domain; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

DR SMART: SM00261; FU; 1.

09W6F6	gallus gall
09BG64	oryctolagus
023821	caenorhabdi
09BG65	oryctolagus
026569	schistosoma
026566	schistosoma
026567	schistosoma
026568	schistosoma
09U5A8	bombyx mori
09NUT5	biotomalari
073798	xenopus lae
093457	scophthalmu
09VD94	drosophila
09VJ04	xenopus lae
088459	mus musculu
063721	rattus norv
063720	rattus norv
09QVW4	mus musculu
09YGH8	scophthalmu
09VJ04	drosophila
03VER2	drosophila
016131	caenorhabdi
09NOK4	sus scrofa
09BH16	anopheles g
088458	mus musculu

Query Match 99.5%; Score 2275; DB 4; Length 419;
Best Local Similarity 99.5%; Pred. No. 4.9e-190;
Matches 417; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELALACRWGLLALLPAGA...VGRGPPDAHVAVNLSRYEG 419
DB 1 MELALACRWGLLALLPAGA...VGRGPPDAHVAVNLSRYEG 419
QY 61 ELTFLPTNASLFLDIOEGVGLIAHNGVRLRIVRGCTOLFEDNYALAVLDNG 120

```
|||||
Db 61 ELTYLPTNANSLFLODIOEVGVYVLIANOVROVPLQRLRIVRGQLEEDNATALAVLNG 120
OY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIIQRPOLCYODTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIIQRPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRIVCAGGACRCKGRLPTDCCHQC 240
Db 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRIVCAGGACRCKGRLPTDCCHQC 240
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
OY 301 YNYLSTDVGSCTLYCPLNNOEVTAEADGTQRCCKSKPCARGTHSLPRPAVPVPLRMQP 360
Db 301 YNYLSTDVGSCTLYCPLNNOEVTAEADGTQRCCKSKPCARGTHSLPRPAVPVPLRMQP 360
OY 361 GFAHVPISFLRPSMDLVSAFYSLPLAPISPVSVGVGPPDAHVAVNLSTRYG 419
Db 361 GFAHVPISFLRPSMDLVSAFYSLPLAPISPVSVGVGPPDAHVAVNLSTRYG 419

RESULT 2
O18735 PRELIMINARY: PRT: 1259 AA.
AC 018735:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ERBB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB008451; BAA23127.1; -.
DR HSSP: P06213; 1IRK.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;
```

```
|||||
Db 121 DPLSGITAPAPAGGRLRELQRLSLTEILKGVLIIQRPOLCHQDTILMKDIFHKNNOLA 180
OY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRIVCAGGACRCKGRLPTDCCHQC 240
Db 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRIVCAGGACRCKGRLPTDCCHQC 240
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
OY 301 YNYLSTDVGSCTLYCPLNNOEVTAEADGTQRCCKSKPCAR----GTHSLPRPAVPV 355
Db 301 YNYLSTDVGSCTLYCPLNNOEVTAEADGTQRCCKSKPCARCYGLGMHLREVAVLSAN 360
OY 356 LRMOPG--PAHVPISFLRPSMDLVSAFYSLPLAP 387
Db 361 IOEFAGCKKIFGSLAFLEPSEDDPASPNTADLP 394

RESULT 3
O90836 PRELIMINARY: PRT: 527 AA.
AC 090836:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EGF/TGF-ALPHA RECEPTOR PRECURSOR.
GN C-ERBB.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92123214; PubMed=1732751;
RA Flickinger T.W., Maibhe N.J., Kung H.J.;
RT "An alternatively processed mRNA from the avian c-erbB gene encodes a
RT soluble, truncated form of the receptor that can block ligand-
RT dependent transformation."
RL Mol. Cell. Biol. 12:883-893(1992).
DR EMBL: M7637; AAA48759.1; -.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 2.
KW Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 527 EGF/TGF-ALPHA RECEPTOR.
SQ SEQUENCE 527 AA; 58353 MW; 764564ABC05258 CRC64;
```

```
Query Match 34.7%; Score 794; DB 13; Length 527;
Best Local Similarity 46.0%; Pred. No. 5,76-61;
Matches 155; Conservative 55; Mismatches 109; Indels 18; Gaps 7;
```



```

OY 244 CTPGPHSCCLACLFHNHSGIEIHELCPALMYNDTFESMNGPARGYFGSCVTACAPYV 303
DB 220 CTCGPHSCCLACRFRFDATCKTCKCPPLVLYNPTTYMDVNPBGKISFGATCYRCPHNY 309
OY 304 LSTDVGSCTVCPPLHNOETVAEDGTQRCCKSCSKPCAR 340
DB 310 VTFDHGSCVRSCTNFDTYEV-EENGVRCRCKKCDGLCSK 345

RESULT 4
OY2795 PRELIMINARY; PRT; 405 AA.
AC 092795; 000732; 000688;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (P60 EPIDERMAL GROWTH
DE FACTOR RECEPTOR).
GN EGF.
OS Homo sapiens (Human).
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Mahle N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RL factor receptor gene encodes a truncated form of the receptor.";
RN Nucleic Acids Res. 24:4050-4056(1996).
RP 12
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RL epidermal growth factor receptor expression in human placenta.";
RN Mol. Reprod. Dev. 41:149-156(1995).
RP 13
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Ilekis J.V., Garlit J., Scoccia B.;
RL Gyn. Onc. 65:0-0(0).
RN 14
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Ilekis J.V.;
RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN 15
RP SEQUENCE FROM N.A.
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downard J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RL expression of the amplified gene in A431 epidermoid carcinoma cells.";
RN Nature 309:418-425(1994).
RP 16
RP SEQUENCE FROM N.A.
RX MEDLINE=85267689; PubMed=2991749;
RA Marinho G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H.,
RA Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A.,
RA Hunts J.H., Shimizu N., Pastan I.;
RT "Structure and localization of genes encoding aberrant and normal
RL cells.";
RN Mol. Cell. Biol. 5:1722-1734(1985).
RP 17
RP SEQUENCE FROM N.A.
RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Schein Sincclair C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,

```

```

RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maille N.J.;
RT *Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EgrR transcripts encoding truncated receptor
RT isoforms";
RL Genomics 0:0-0(2000).
RL (8)
RN SEQUENCE FROM N.A.
RA Reiter J.L., Eley G.D.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U48725; AAC50797.1; -
DR EMBL; U48722; AAC50802.1; -
DR EMBL; U48724; AAC50796.1; -
DR EMBL; U48726; AAC50798.1; -
DR EMBL; U48723; AAC50803.1; -
DR EMBL; U95089; AAB53063.1; -
DR EMBL; AF288738; AAC53787.1; -
DR InterPro: IPR000494; EGRF.L.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L-domain; 1.
DR SMART; SM00261; FU; 1.
KW Signal; Receptor.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 405 EPIDERMAL GROWTH FACTOR RECEPTOR.
SQ SEQUENCE 405 AA; 44664 MW; F5DEB31787EF1822 CRC64;

Query Match 34.7%; Score 793; DB 4; Length 405;
Best Local Similarity 45.3%; Pred. No. 5,1e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4

QY 11 LLLALLPRGAA--STOVCTGTDMLRLPASPETHDMLRLHLYOGCVOVGNLELYLPYN 68
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 14 LLAALCPASRALEEKVKCCGQSTNKLTLQGTGFEDHFLSLQRMFNCCVVLGNLEITYQGN 73

QY 69 ASLSFLDIOIEVQGVYLLAHNOVROYPLRLKIVKSTOLFEDNYTALAVLDNGDPLNNTTP 128
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 74 YDLSEFLKTYIEVAGVYVLLALNVTVERPLENLOIIRGNMYEYSYALAVLSNYD----- 126

QY 129 VTGASPGGLREVLQSLPEFLKGGVLIQNPOLCYQDPTLMDIFPKNNQALLTLIDTR 188
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 127 ---AKTKGTKEPLPMNLOELHGAVFESNNPCLAVESIQMDVIYSSDPLSNMSDPQNH 183

QY 189 SRACSPSPMCKSGKRCMGSSSEDCSLRTFYAGGCA-FCKGRLPTDCGHEQCAAGCTGP 247
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 184 LGSCKCKPSCSCNGSCMGAGEKNCQKLTIKLQAQCSGKCRKRSYSDCHNCCAACTGP 243

QY 248 KHSDCIACLFHNHSGICELHCPALVTYNTDFESMPNEGRYTFGASCVTACPYNYLSTD 307
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 244 RESDLCVQRKFDEATCKDQTCPLMLNPTTYQMDVNPFGKYSFGATCVKCKPRNVVTD 303

QY 308 VGSCTLVCPRLHNOEYTAEDGTORCKCSKPAR 340
   ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
DB 304 HGSCVRACGADSYEM-EDDGVKCKCKCEGPKR 335

RESULT 5
Q9H2C9 PRELIMINARY; PRT; 628 AA.
AC Q9H2C9;
DT 01-MAR-2001 (TREMBLrel. 16; Created)
DT 01-MAR-2001 (TREMBLrel. 16; Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17; Last annotation update)
DE TRUNCATED EPIDERMAL GROWTH FACTOR RECEPTOR.
GN EGRF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homalidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84219729; PubMed=6328312;

```

RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.,
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.",
 RL Nature 309:418-425(1984).
 [2]
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE=85267689; PubMed=2991749;
 RA Merlino G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H.,
 RA Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A.,
 RA Hunts J.H., Shimizu N., Pastan I.,
 RT "Structure and localization of genes encoding aberrant and normal
 RT epidermal growth factor receptor RNAs from A431 human carcinoma
 RT cells.",
 RL Mol. Cell. Biol. 5:1722-1734(1985).
 [3]
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Mähle N.J.,
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.",
 RL Nucleic Acids Res. 24:4050-4056(1996).
 [4]
 RN RP SEQUENCE FROM N.A.
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Schell Sinclair C., Pearce R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balsubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Mähle N.J.,
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.",
 RL Genomics 0:0-0(2000).
 [5]
 RN RP SEQUENCE FROM N.A.
 RA Reiter J.L., Eley G.D.,
 RL Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF288738; AAC35790.1;
 DR InterPro: IPR000345; CytC_heme_bind.
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 4.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 KW Receptor.
 SO SEQUENCE 628 AA; 69228 MW; 3A00A5511A3B6AE2 CRC64;

Query Match 34.7%; Score 793; DB 4; Length 628;
 Best local similarity 45.3%; Pred. No. 8,5e-61;
 Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;
 QY 11 LLLALLPPGAA--STOYCTGTDKRLIPASPEPHLDMLRLHYGCCOYVGNLELTPLTN 68
 DB 14 LLLALCPASALEEKRYCOGTSTNKLTLQGFEDHFLSTQFMFNVCVLENNETTYVORN 73
 QY 69 ASLSPLODDIOEGYVLIANOVROYPLORLIVRGTOLEFNNVAVALVDNGPLNNTTP 128
 DB 74 YDLSFLTTIDEVAGYVLIANTVERIPLENLQIRGMVENSVALAVLSNTP----- 126
 QY 129 VTGASPGGLELRLSTLEILKGVLIQRPOLCYODTILMKDIFHKNNOLATLIDTNR 188
 DB 127 ---ANKTGKLELPMRNQELHGAVERSNPALCNVESIOWRIYSDPLSNMSMFOHN 183
 QY 189 SRACHPCSPKCKSRCKGSESDQSLTRTYCAGGCA-RCKGPLPLDCCHEGCAAGCTG 247
 DB 184 LGCSCQKCDPCSPNGSCGAGAEENCKLTKLTICAQCGSGRCGRKSPDCHNCCAACTGP 243
 QY 248 KHSDDCLACHFNNSGICELHCPALMYNTDPSFMPRPEGRYTFGASCYACAYNTLSTD 307
 DB 244 RESDCLVCRFRDPAVTCRDTCPPLMLYNPTTYQMDVNPBGKYSFGATCVKCPRNTVYTD 303

OY 308 VGSCTVCPHLNHOEYTAEDGTORCEKSCSKPCAR 340
 DB 304 HGSCVRACGADSYEM-EDDGVKCKCKGCPCK 335
 RESULT 6
 Q14226
 ID 014226 PRELIMINARY; PRT; 657 AA.
 AC 014226; 093010;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EPIDERMAL GROWTH FACTOR RECEPTOR (A431-SPECIFIC P115 EPIDERMAL GROWTH
 DE FACTOR RECEPTOR).
 GN EGFR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE=85267689; PubMed=2991749;
 RA Merlino G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H.,
 RA Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A.,
 RA Hunts J.H., Shimizu N., Pastan I.,
 RT "Structure and localization of genes encoding aberrant and normal
 RT epidermal growth factor receptor RNAs from A431 human carcinoma
 RT cells.",
 RL Mol. Cell. Biol. 5:1722-1734(1985).
 [2]
 RN RP SEQUENCE OF 501-544 FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Mähle N.J.,
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.",
 RL Nucleic Acids Res. 24:4050-4056(1996).
 [3]
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.,
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.",
 RL Nature 309:418-425(1984).
 [4]
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE=85267689; PubMed=2991749;
 RA Merlino G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H.,
 RA Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A.,
 RA Hunts J.H., Shimizu N., Pastan I.,
 RT "Structure and localization of genes encoding aberrant and normal
 RT epidermal growth factor receptor RNAs from A431 human carcinoma
 RT cells.",
 RL Mol. Cell. Biol. 5:1722-1734(1985).
 [5]
 RN RP SEQUENCE FROM N.A.
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Schell Sinclair C., Pearce R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balsubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Mähle N.J.,
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.",
 RL Genomics 0:0-0(2000).
 [6]
 RN RP SEQUENCE FROM N.A.
 RA Reiter J.L., Eley G.D.,
 RL Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; K03193; AAA52371.1;
 DR EMBL; U48728; AAC50800.1;
 DR EMBL; AF288738; AAC35788.1;

DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 3.
 KW Receptor.
 SQ SEQUENCE 657 AA; 72395 MW; CEC1D6C7C4C52C2A CRC64;

Query Match 34.7%; Score 793; DB 4; Length 657;
 Best Local Similarity 45.3%; Pred. No. 8,9e-61;

Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

QY 11 LTLALLPFGAA--STOVCGTDMKRLPASPETHDMLRHLVGGCGVGVGNLELYPTN 68
 DB 14 LLAALCPASRALEEKVKVCGTGNKLTQLGTFEDHFLSLORMNCEVYVGNLEITYVOR 73
 QY 69 ASLSFLADIQEVGVYVLAHNOVROVPLQRLIRVGTQLFEDNYALAVLNDGDLNNTTP 128
 DB 74 YLSEFLKTIQEVGVYVLAHNOVROVPLQRLIRVGTQLFEDNYALAVLNDGDLNNTTP 126
 QY 129 VVGASPGGLREQLSLTFLILKGVLIQNPOLCYODTILMKDIFRKNQALATLIDTR 188
 DB 127 ---ANKTGKELPMRNLQIILGAVRFSNNPALCNVESIOWMDIVSSDPLSNMSMDPOMH 183
 QY 189 SRACHPCSPMGKSGRCWGESSEDCSLFRVCAGGA-RCKGPIPTDCHEOCAGCTGP 247
 DB 184 LGSCKQDPCSPGSCWGEENGCKLTKITICAOQSGRCRKSDDCHNOCAAGCTGP 243
 QY 248 KHSDELACLFHNHSGICELCPALVTYNTDFESHPNDEGRYTFGASCVTACPYNYLSTD 307
 DB 244 RESDCLVCRKFRDEATCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCYKCPRYVYTD 303
 QY 308 VGSCTLVCPHNOEYTAEDGTORCEKSPCAR 340
 DB 304 HGSCVRACGADSYEM-EEDGVKCKKCEGPCRK 335

RESULT 7
 Q9GZXI PRELIMINARY; PRT; 705 AA.
 AC Q9GZXI.
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE TRUNCATED EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (P110 EPIDERMAL GROWTH FACTOR RECEPTOR).
 GN EGFR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schehl C.M., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maibale N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the extracellular domain of the receptor.";
 RL Submitted (Feb-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W., Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J., Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85267689; PubMed=2991749;

RA Merlino G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H., Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A., Hunts J.H., Shimizu N., Pastan I.;
 RT "Structure and localization of genes encoding aberrant and normal epidermal growth factor receptor RNAs from A431 human carcinoma cells.";
 RL Mol. Cell. Biol. 5:1722-1734(1985).
 RN [4]

RP SEQUENCE FROM N.A.
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Maibale N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [5]

RP SEQUENCE FROM N.A.
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Schehl Sinclair C., Pearall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maibale N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and mouse alternative EGFR transcripts encoding truncated receptor isoforms.";
 RL Genomics 0:0-0(2000).
 RN [6]

RP Reiter J.L., Eley G.D.;
 RL Submitted (Jul-2000) to the EMBL/Genbank/DBJ databases.
 RA EMBL: AF15253; AAC3240.1;
 DR EMBL: AF288738; AAC35786.1;
 DR InterPro: IPR000345; CYC_heme_bind.
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 4.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 KW Signal; Receptor.
 FT SIGNAL 1 24
 FT CHAIN 25 705
 FT POTENTIAL.
 FT TRUNCATED EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT SEQUENCE 705 AA; 77312 MW; ACEF14942FF1650C CRC64;

Query Match 34.7%; Score 793; DB 4; Length 705;
 Best Local Similarity 45.3%; Pred. No. 9,7e-61;

Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

QY 11 LTLALLPFGAA--STOVCGTDMKRLPASPETHDMLRHLVGGCGVGVGNLELYPTN 68
 DB 14 LLAALCPASRALEEKVKVCGTGNKLTQLGTFEDHFLSLORMNCEVYVGNLEITYVOR 73
 QY 69 ASLSFLADIQEVGVYVLAHNOVROVPLQRLIRVGTQLFEDNYALAVLNDGDLNNTTP 128
 DB 74 YLSEFLKTIQEVGVYVLAHNOVROVPLQRLIRVGTQLFEDNYALAVLNDGDLNNTTP 126
 QY 129 VVGASPGGLREQLSLTFLILKGVLIQNPOLCYODTILMKDIFRKNQALATLIDTR 188
 DB 127 ---ANKTGKELPMRNLQIILGAVRFSNNPALCNVESIOWMDIVSSDPLSNMSMDPOMH 183
 QY 189 SRACHPCSPMGKSGRCWGESSEDCSLFRVCAGGA-RCKGPIPTDCHEOCAGCTGP 247
 DB 184 LGSCKQDPCSPGSCWGEENGCKLTKITICAOQSGRCRKSDDCHNOCAAGCTGP 243
 QY 248 KHSDELACLFHNHSGICELCPALVTYNTDFESHPNDEGRYTFGASCVTACPYNYLSTD 307
 DB 244 RESDCLVCRKFRDEATCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCYKCPRYVYTD 303
 QY 308 VGSCTLVCPHNOEYTAEDGTORCEKSPCAR 340
 DB 304 HGSCVRACGADSYEM-EEDGVKCKKCEGPCRK 335

```

RESULT      8
QSERV6      PRELIMINARY;      PRT:      643 AA.
ID           01-MAR-2001 (TREMBlrel. 16, Created)
AC           01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT           01-MAR-2001 (TREMBlrel. 17, Last annotation update)
DE           01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE           EPIDERMAL GROWTH FACTOR RECEPTOR ISOFORM 2.
GN           EGFR.
OS           Mus musculus (Mouse).
OC           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC           NCBI_TaxID=10090;
OX           [1]
RN           SEQUENCE FROM N.A.
RC           STRAIN=C3H/101, 129/SVJ, 129/SVEV/TAC;
RA           Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA           Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA           Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA           Maibhe N.J.;
RT           *Comparative genomic sequence analysis and isolation of human and
RT           mouse alternative Egfr transcripts encoding truncated receptor
RT           isoforms.";
RL           Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR           EMBL; AF275366; AAG28046.1; -
DR           EMBL; AF275364; AAG28046.1; JOINED.
DR           EMBL; AF275365; AAG28046.1; JOINED.
DR           MGI:95294; Egfr.
DR           InterPro: IPR000345; Cytc_heme_bind.
DR           InterPro: IPR000494; EGFR_L.
DR           InterPro: IPR002174; Furin-like.
DR           Pfam: PF00757; Furin-like; 1.
DR           Pfam: PF01030; Recep_L_domain; 2.
DR           SMART: SM00261; FU; 4.
DR           PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
KW           Receptor.
SQ           SEQUENCE      643 AA: 71477 MW; DEF22002C84911B1 CRC64;

Query Match      34.5%; Score 789; DB 11; Length 643;
Best Local Similarity 46.3%; Pred. No. 1,9e-60;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;

OY      11 LLLALLPGGA--STQVCTGTDMKRLRLPASPETHLDMLRHLYGCGVQVGNLELYLPTN 68
DB      14 LITLCAAGAALEKKVCGGTNRILQLGTFEDHFLSLQRMVNNCEVYLGNIETIYVORN 73
OY      69 ASLSEFLQDIQEVQVYLIAHNOVROVPLORLRLVKTQLEFDNYKALAVLDNCDPLANTTP 128
DB      74 YDLSEFLKTIQEVAGVYLIANTVERIPLENIQIRGNALYENTYALALISN----- 124
OY      129 VNGASPGGLRELOLRTEILKGVLIQIRNPOLCYQDTILMKDI---FHKNNQIALTLI 184
DB      125 -YGTNATGRLRLPRLMRLQELILGAVFNSNNPILCNMDITQMDIYQNVMSNMSKL---- 180
OY      185 DYNRSRACHPCSPMKSGRSCWGESSESDCSLRTVYAGGCA-RCKGPLPTDCHEQCAAG 243
DB      181 -QSHPSCKPCPSKPCSCNGSCWGEENCKLTKIICAOQCSHRCGRSPSCCHNCAG 239
OY      244 CTGPKHSDCLACILHNHSGICELHCPALTYTNTDFESHPNPGRTFGASCTYACPYV 303
DB      240 CTGPRSDCLVQCKFQDEATCKDTCPLMLYNFTTYQMVNPEKYSFGATCVKCPRYN 299
OY      304 LSTDVSGCTLVCPILHNOEYTAEDGTORCEKCSPCAR 340
DB      300 VVTDGSCVACGPDYEV-EEDGIRKCKKCGDGPCKR 335

```

```

DT      01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT      01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE      EPIDERMAL GROWTH FACTOR RECEPTOR (EPIDERMAL GROWTH FACTOR RECEPTOR
DE      ISOFORM 3).
GN           EGFR.
OS           Mus musculus (Mouse).
OC           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC           NCBI_TaxID=10090;
OX           [1]
RN           SEQUENCE FROM N.A.
RC           STRAIN=C57BL/6J; TISSUE=LIVER.
RA           Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,
RA           Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA           Maibhe N.J.;
RT           *Alternative transcripts from the Human and Mouse EGFR Genes Encode
RT           Carboxy-Terminal Truncated Receptors.";
RL           Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN           [2]
RP           SEQUENCE FROM N.A.
RC           STRAIN=C3H/101, 129/SVJ, AND 129/SVEV/TAC;
RA           Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA           Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA           Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA           Maibhe N.J.;
RT           *Comparative genomic sequence analysis and isolation of human and
RT           mouse alternative Egfr transcripts encoding truncated receptor
RT           isoforms.";
RL           Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN           [3]
RP           SEQUENCE FROM N.A.
RC           STRAIN=C57BL/6J; TISSUE=LIVER.
RX           MEDLINE=21085660; Pubmed=11217851;
RA           Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA           Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA           Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA           Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA           Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA           Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA           Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA           Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA           Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA           Blake J., Boffelli D., Fujuno M., Carninci P., de Bonaldo M.F.,
RA           Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA           Gustincich S., Hill D., Hofmann M., Hume D.A., Kanliya M., Lee N.H.,
RA           Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombetris P.,
RA           Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA           Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA           Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA           Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA           Hayashizaki Y.;
RT           "Functional annotation of a full-length mouse cDNA collection.";
RL           Nature 409:685-690(2001).
DR           EMBL; AF124513; AAD44149.1; -
DR           EMBL; AF275366; AAG28047.1; -
DR           EMBL; AF275364; AAG28047.1; JOINED.
DR           EMBL; AF275365; AAG28047.1; JOINED.
DR           EMBL; AK004944; BAB23688.1; -
DR           EMBL; AK004883; BAB23641.1; -
DR           EMBL; AK004911; BAB23662.1; -
DR           MGI:95294; Egfr.
DR           InterPro: IPR000494; EGFR_L.
DR           InterPro: IPR002174; Furin-like.
DR           Pfam: PF00757; Furin-like; 1.
DR           Pfam: PF01030; Recep_L_domain; 2.
DR           SMART: SM00261; FU; 3.
KW           Receptor.
SQ           SEQUENCE      655 AA: 72906 MW; 6B34063B1BC928CB CRC64;

Query Match      34.5%; Score 789; DB 11; Length 655;
Best Local Similarity 46.3%; Pred. No. 2e-60;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;

```

```

QY 11 LLLALLPPGAA--STGYCTGDMKRLRLPASPETHLMDLRHLHYGCGVYVGNLELTPTN 68
  || || || || || || || || || || || || || || || || || || || || || ||
Db 14 LITLCAAGGALAEKKVCGGTSNRLTQGTFFEDHFLSLQRMNNEVYVGNLEITYVGRN 73
QY 69 ASLSFTLDIOEVGYVLLAHNOVROVPLQRLRYRGTOLEFEDNVALAVLDGDPPLNNTTP 128
  || || || || || || || || || || || || || || || || || || || || || ||
Db 74 YDLSFLKTIQEVAGVYLLALMTVERIPLENQIIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDI----FHKNNOLATLI 184
  || || || || || || || || || || || || || || || || || || || || || ||
Db 125 -YGTNRKTGLRELPMRNLOEILLGAVRFNSNPILCMNDTIQMRDIQVNFMSMSMDL--- 180
QY 185 DTNRSRACHPCSPMKGSRGCGSESEDCOSLRTVYACAGCA-RCKGPLPTDCCHEQCAAG 243
  || || || || || || || || || || || || || || || || || || || || || ||
Db 181 -QSHPSKPCDPCSPNCGSGWGGENCQKLTIIICAQGCCHRGSRSPDCHNOQCAAG 239
QY 244 CTGPKHSDCLALIFHNHSGICELHCPALVTYNTDFFESMPNPEGRTYTGASCYVACPYNY 303
  || || || || || || || || || || || || || || || || || || || || || ||
Db 240 CTGPRESDCLVQKRFQDEATKDCDPCPLMLNPTTYQMDVNPBEGKYSFGATCVKCKPRNY 299
QY 304 LSTDVGSCTVCPPLHNOEVTAEADGTORCEKSKPCAR 340
  || || || || || || || || || || || || || || || || || || || || || ||
Db 300 VVTDHSGCVACGPDYEV-EEDGIRKCKKCDGCRK 335

RESULT 10
Q9EP98 PRELIMINARY; PRT; 1210 AA.
ID Q9EP98;
AC Q9EP98;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR ISOFORM 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, 129/SVEVTAJ;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF275366; AAG28045.1; JOINED.
DR EMBL: AF275365; AAG28045.1; JOINED.
DR EMBL: AF275367; AAG24386.1; -.
DR MGD: MGI:95294; Egr.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_PKinase.
DR InterPro: IPR002174; Euk_PKinase.
DR InterPro: IPR002290; Ser_Thr_Kin_actsite.
DR InterPro: IPR001245; Tyr_Kin.
DR Pfam: PF00757; Furin-1like; 1.
DR Pfam: PF00069; pkinase; 1.

```

```

DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00261; FU; 5.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00219; TYKIC; 1.
DR PROSITE: PS00107; PROTEIN KINASE_ATP; 1.
DR PROSITE: PS00111; PROTEIN KINASE_DOM; 1.
DR ATP-binding; Receptor; Transferase.
KW SEQUENCE 1210 AA; 134841 MW; 62CD021C9DE32E18 CRC64;

Query Match 34.5%; Score 789; DB 11; Length 1210;
Best Local Similarity 46.3%; Pred. No. 4e-60;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;

QY 11 LLLALLPPGAA--STGYCTGDMKRLRLPASPETHLMDLRHLHYGCGVYVGNLELTPTN 68
  || || || || || || || || || || || || || || || || || || || || || ||
Db 14 LITLCAAGGALAEKKVCGGTSNRLTQGTFFEDHFLSLQRMNNEVYVGNLEITYVGRN 73
QY 69 ASLSFTLDIOEVGYVLLAHNOVROVPLQRLRYRGTOLEFEDNVALAVLDGDPPLNNTTP 128
  || || || || || || || || || || || || || || || || || || || || || ||
Db 74 YDLSFLKTIQEVAGVYLLALMTVERIPLENQIIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDI----FHKNNOLATLI 184
  || || || || || || || || || || || || || || || || || || || || || ||
Db 125 -YGTNRKTGLRELPMRNLOEILLGAVRFNSNPILCMNDTIQMRDIQVNFMSMSMDL--- 180
QY 185 DTNRSRACHPCSPMKGSRGCGSESEDCOSLRTVYACAGCA-RCKGPLPTDCCHEQCAAG 243
  || || || || || || || || || || || || || || || || || || || || || ||
Db 181 -QSHPSKPCDPCSPNCGSGWGGENCQKLTIIICAQGCCHRGSRSPDCHNOQCAAG 239
QY 244 CTGPKHSDCLALIFHNHSGICELHCPALVTYNTDFFESMPNPEGRTYTGASCYVACPYNY 303
  || || || || || || || || || || || || || || || || || || || || || ||
Db 240 CTGPRESDCLVQKRFQDEATKDCDPCPLMLNPTTYQMDVNPBEGKYSFGATCVKCKPRNY 299
QY 304 LSTDVGSCTVCPPLHNOEVTAEADGTORCEKSKPCAR 340
  || || || || || || || || || || || || || || || || || || || || || ||
Db 300 VVTDHSGCVACGPDYEV-EEDGIRKCKKCDGCRK 335

RESULT 11
Q9OX70 PRELIMINARY; PRT; 1209 AA.
ID Q9OX70;
AC Q9OX70;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA MEDLINE=90256888; PubMed=2342466.
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue."
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Guttridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
EMBL: M37394; AAF14008.1; -.

```

DR HSSP: P06239; 3LCK.
DR InterPro: IPR0000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SMO0261; FU; 3.
DR SMART: SMO0219; TYRCK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase
SO SEQUENCE 1209 AA; 134890 MW; 96FEF76CC1B7773 CRC64;

Query Match:	34.2%	Score 781.5;	DB 11;	Length 1209;
Best Local Similarity	46.0%;	Pred. No. 1.8e-59;		
Matches 157;	Conservative 44;	Mismatches 117;	Indels 23;	Gaps 7;

Qy	3	LAALRMELTALLALPPGA-ASTOVCTGDMKLRIPASEFELDLMLRHYOOCVQVONLE	61
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	15	LAALCAAG-----GALREKKYQCGGSRNKLQTQGFEDHFELSLOQMFNNCEVYVNL	66
Qy	62	LTLYETPNASLSFLODIOEVGVGLIAHQVAPLQRLRIYRGTLQTFEDNYALAVLNDG	121
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	67	ITYVQRNADISFLKTIQVAGVYVIALNTVRIPIEMQIIRGALVENVYALVALVSN--	124
Qy	122	PLMNTTPVTGASPGGLRELOLRSLTEILKGGVLLIQRNPQLCYOPTYLMKDIFFHKNNQAL	184
		: : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	125	-----YGNKIGLRLLPRRNIOELLIGAVRRSNMPLDNCMTIQMRDVI-QDVFSLN	175
Qy	132	TLIDITNRS-RACHPCSPMCKSRCKWGESSEDCQSLTRVYACAGCA-RCKGPLPDCCHEQ	238
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	176	MSMDVQRRLTGCPCDPKPCSPGSGCWGRREENCKRLTKICAQOCSRRRCRGHSPSDCCHQ	235
Qy	240	CAAGCTGKRHSCLACLTFHNSGICELCPALVYNTDFESMNPGRGARYFGASGYTAC	289
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	236	CAAGCTGTRSDCLCHFRDEACCKDTCPLMLYNTPTYQMDVNPBSKTSFGATCYKKC	295
Qy	300	PNYNTSDVSGCTLVGLPHNDVYAEADSTQCEKCSKPCAR	340
		: : : : : : : : : : : : : : : : : : : : :	
Db	236	PNRYVTVTHGSCVCRACGPDYEV--EEDVSYSCKKCKDGGCRK	335

RESULT	12		
0922N7			
ID	0922N7	PRELIMINARY;	PRT: 1308 AA.
AC	0922N7		
DT	01-MAY-1999 (TReMBLrel. 10, Created)		
DT	01-MAY-1999 (TReMBLrel. 10, last sequence update)		
DT	01-JUN-2001 (TReMBLrel. 11, last annotation update)		
DE	RECEPTOR TYROSINE KINASE.		
GN	ERBB4.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;		
CC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=HEART.		
RA	Zhao Y.Y., Kelly R.A., Smith T.W.;		
RL	Submitted (Jan-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF041838; AAD08899.1; -		
DR	HSSP; P06213; IIRK.		
DR	InterPro; IPR000494; EGFR_L.		
DR	InterPro; IPR000719; Euk_kinase.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR001368; TNFR_c6.		
DR	InterPro; IPR001245; Tyr_Kin.		
DR	Pfam; PF00757; Furin-like; 1.		

DR Pfam: PF00069; kinase: 1.
DR Pfam: PF01030; Recep_L_domain: 2.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SMO0261; FU: 4.
DR SMART: SMO0219; Tyrc: 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP: 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM: 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR: 1.
DR PROSITE: PS00652; TNFR_NGFR1; UNKNOWN: 1.
KW ATP-binding; kinase; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1308 AA; 146957 MW; D944BBD0996408B41 CRC64;

Query Match	34.1%	Score 779.5	DB 11	Length 1308
Best Local Similarity	45.98%	Pred. No. 2.9e-59		
Matches 163, Conservative	45	Mismatches 126	Indels 21	Gaps 8

```

OY      1  MELA-AICRMILLI- -ALLPFGASATVCSTGDKLILPASPENHIMDLNHLKQSCVVO  57
Db      1  MKLATGMLWMSLLVAARTVQPSASQSVCASTENKLSLSLDEQYHALKRYENCEVVM  60

OY      58  GNLEITLYPTASISFLDIOEVQGYLLAHNOYROYPLDRLRVSTOLEFEDNYALAVL  117
Db      61  GNLEITLYSTENHRDLSFLKSIIEVGYVALVAINOFRYLPLELRILNRTKLYEERYALAI  120

OY      118  DNGDELNTTTPVTGASPGELRELOLRSTELLKSGVLIOIRNPOLCYODTLIMKDIFFKNN  177
Db      121  LNYRKDNF-----GIQELGLKNTLELLNGVYVDQKKFLCYADTLIHMDIVRNPW  171

OY      178  QALATLIDTNSRACHPCSPMCKSRCKWGESSEDCQSLTRYVACGC- -AACKPLPLPDC  236
Db      172  PSMNTLVSTISSCGCRHKSCSTG-RCMGPRPEHNCOTLLTRVCACQCDGRCYGGYVSDC  230

OY      237  HEQCAAGCTGGRHSACLACLFHNSHGICELCPALVYNTDFESMNPBGRATYFGASCV  266
Db      231  HRECAAGCGSGRKDDCFACMNFNDGACVYTCQPTFLVYNPTTQLDEHNPAAKYTYGAFCV  290

OY      297  TACPRNTYSTVDSGCTLYCPLPHNOEYVAEDSTQCEKSCFKCAR- -GTHSLI  346
Db      291  KCKPNEF- -VSSSCSVRACPSSSKMEV- -ENKIKCKCPCTDLCPCACGIGTGSILM  343

```

SEQUENCE	RESULT	13
Q9ESEO	Q9ESEO	PRELIMINARY; PRT; 473 AA.
AC	Q9ESEO.	
DT	01-MAR-2001 (Tremblrel, 16, Created)	
DT	01-MAR-2001 (Tremblrel, 16, Last sequence update)	
DT	01-JUN-2001 (Tremblrel, 17, Last annotation update)	
DE	EPIDERMAL GROWTH FACTOR RECEPTOR RELATED PROTEIN.	
GN	ERRP.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Euhetia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=GASTRO-DUODENAL, MUCOUS;	
RA	Yu Y., Moshier J.A., Majumdar A.P.N.;	
RT	"Cloning of a novel EGFR-related peptide: A putative negative	
RL	regulator of EGFR.";	
DR	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: AF187818; AAC17037.1; -	
DR	InterPro: IPR000494; EGFR_L.	
DR	InterPro: IPR002174; Furin-like.	
DR	Pfam: PF00757; Furin-like; 1.	
DR	Pfam: PF01030; Recep_L_domain; 2.	
DR	SMART: SM00261; Fu; 1.	
KW	Receptor.	
SO	SEQUENCE	473 AA; 52903 MW; 4E0D7C3074F28973 CRC64;

Query match 33.8%; Score 773; DB 11; Length 473;

Best Local Similarity 45.7%; Pred. No. 3.4e-59;
Matches 154; Conservative 43; Mismatches 118; Indels 22; Gaps 6;

QY 11 LLLALLPPGAA--STOVCTGDMKRLRSPETHLMDLRLHYOGCVOVGNLELYLPIN 68
DB 14 LLLALLPPGAA--STOVCTGDMKRLRSPETHLMDLRLHYOGCVOVGNLELYLPIN 73
QY 69 ASLSFLODIOEVGYVLIHNOVROVPLORLRIVRGTOLEFENYALAVLDNGDPLNNTPT 128
DB 74 YDLSTLKTIOEVAGYFLIANTVERIPSEDLOIRGNALYEMTYALALISN----- 124
QY 129 VTGASPGILRELQSLRELEIKGVLIORNPOLCYODTILMKDI---FHKNNOLATLTI 184
DB 125 -GTRNRTGRELPMRNLDEILIGAVRFSNNPILCNMDITQWRDIQVNVFMSMSMDL--- 180
QY 185 DNRBRACHPCSPMCKSGMCESESDCSLRTVACAGCA-RCKGRLPTDCCHEQCAAG 243
DB 181 -OSHSSCPKCPSCPNCSGCGEENCQKIRKICAOCSHRCRSPSDCCCHNOCAAG 239
QY 244 CTGPKHSDCLACIHFHNSGICELHCPALVYNTDFESMPNREGRTFGASCVTACPNY 303
DB 240 CTGPKHSDCLACIHFHNSGICELHCPALVYNTDFESMPNREGRTFGASCVTACPNY 299
QY 304 ISTDVSGCTLVCPRLHNOEYTAEDGTORCEKCSKPCAR 340
DB 300 VYTDHSGCVACGPDYEV-EEDGIRKCKKCDGPKRK 335

RESULT 14
Q9BUD7 PRELIMINARY; PRT: 331 AA.
AC Q9BUD7;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILAR TO V-ERB-B2 AVIAN ERYTHROBLASTIC LEUKEMIA VIRAL ONCOGENE
DE HOMOLOG 3.
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=CHORIOCARINOMA;
RA Strausberg R.;
RL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC002706; AAI02706.1;--
SQ SEQUENCE 331 AA: 36489 MW: 4588E8E683FE7E8 CRC64;

Query Match 32.1%; Score 734; DB 4; Length 331;
Best Local Similarity 44.0%; Pred. No. 5.6e-56;
Matches 147; Conservative 49; Mismatches 120; Indels 18; Gaps 8;

QY 10 GLLALLPPGAA--STOVCTGDMKRLRSPETHLMDLRLHYOGCVOVGNLELYLPIN 67
DB 11 GLLFSLANGSEVNSQAVCPGLNLSVTDGAENQYOTLYKIERCEVYMGNLLEYLICH 70
QY 68 NASLSFLODIOEVGYVLIHNOVROVPLORLRIVRGTOLEFENYALAVLDNGDPLNNT 127
DB 71 NADLSFLOMIREVTGYVLIHNOVROVPLORLRIVRGTOLEFENYALAVLDNGDPLNNT 125
QY 128 PYTGASPGILRELQSLRELEIKGVLIORNPOLCYODTILMKDI---FHKNNOLATLTI 187
DB 126 ---NSSHALROLRLQLELISLGGVILEKNDKCHMDPTIDMRDIDRD---AEIVYKD 178
QY 188 RSRACHPCSPMCKSGMCESESDCSLRTVACAGC-ARCKGRLPTDCCHEQCAAGCTG 246
DB 179 NGRSCRPCHGVKRG-RCMNGPSEDCOTILKTICAPQCNCHGCPNPNQCHDECAAGCSG 237
QY 247 FKHSDCLACIHFHNSGICELHCPALVYNTDFESMPNREGRTFGASCVTACPNYLIST 306
DB 238 PDFTDCAFACRHRHNDGACVPCPPLVYVKNLTFQLEPNPHTKYQYGVAVASCPHNFV-V 296

QY 307 DVGSCITVCPILHNOEYTAEDGTORCEKCSKPCAR 340
DB 297 DOTSCYRACPPDKMEVD-KNGLKCEPCGGLCPK 329

RESULT 15
Q9B66 PRELIMINARY; PRT: 149 AA.
AC Q9B66;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE RECEPTOR TYROSINE KINASE ERBB2 (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid:9986;
RN (1)
RP SEQUENCE FROM N.A.
RA Klonisch T., Wolf P., Hombach-Klonisch S., Vogt S., Kuechenhoff A.,
RA Telens F., Fischer B.;
RT "ErbB genes and epidermal growth factor- (EGF-) like ligands in the
RT per-implantation rabbit uterus and blastocyst."
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF333178; AAK14371.1;--
KW Kinase.
FT NON-TER 1 1
FT NON-TER 149 149
SQ SEQUENCE 149 AA: 16240 MW: 7CB3792A54FC49BA CRC64;

Query Match 31.6%; Score 723; DB 6; Length 149;
Best Local Similarity 81.2%; Pred. No. 2e-55;
Matches 121; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 158 NPOLCYODTILMKDI---FHKNNOLATLTIPTNRSRACHPCSPMCKSGMCESESDCSLTR 217
DB 1 NQPCYODTILMKDI---FHKNNOLATLTIPTNRSRACHPCSPMCKSGMCESESDCSLTR 60
QY 218 TVCAGGACARCKRLPTDCCHEQCAAGCTGPKHSDCLACIHFHNSGICELHCPALVYNTD 277
DB 61 TICAGGACARCKQLPTDCCHEQCAAGCTGPKHSDCLACIHFHNSGICELHCPALVYNTD 120
QY 278 TPESMPNREGRTFGASCVTACPNYLIST 306
DB 121 TPESMPNREGRTFGASCVTACPNYLIST 149

Search completed: April 11, 2002, 09:35:08
Job time: 325 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:29:35 ; Search time 20.27 Seconds
(without alignments)
465.164 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287
Sequence: 1 MELALCRWLLALLLPFGA.....VGRGPDPAHVAVLSRYEG 419

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	82.1	782	2	US-09-146-283-4
2	1878	82.1	782	3	US-08-579-823A-4
3	1878	82.1	782	4	US-09-344-195-4
4	1878	82.1	1255	1	US-08-467-083-68
5	1878	82.1	1255	1	US-08-414-417B-68
6	1878	82.1	1255	2	US-08-484-438-8
7	1878	82.1	1255	2	US-08-486-348A-68
8	1878	82.1	1255	2	US-08-625-101-2
9	1878	82.1	1255	2	US-08-468-545B-68
10	1878	82.1	1255	2	US-08-356-786-2
11	1878	82.1	1255	3	US-08-466-680B-68
12	1769	77.4	624	3	US-08-422-108-1
13	793	34.7	644	1	US-08-336-708A-9
14	793	34.7	1210	2	US-08-484-438-7
15	793	34.7	1210	2	US-08-475-035-4
16	775	33.9	911	2	US-08-484-438-10
17	775	33.9	1058	2	US-08-484-438-4
18	775	33.9	1308	2	US-08-484-438-2
19	735.5	32.2	1342	1	US-07-978-895-4
20	735.5	32.2	1342	2	US-08-484-438-9
21	735.5	32.2	1342	2	US-08-473-119-4
22	735.5	32.2	1342	2	US-08-475-352-4
23	734	32.1	1343	6	5183884-4
24	493	21.6	97	1	US-08-421-356-3
25	264.5	11.6	1382	2	US-08-737-715-2
26	257.5	11.3	516	3	US-08-746-559A-4
27	257.5	11.3	1367	2	US-08-249-687C-2

28	257.5	11.3	1367	2	US-08-625-819-2	Sequence 2, Appl
29	257.5	11.3	1367	3	US-08-746-559A-2	Sequence 2, Appl
30	241.5	10.6	486	3	US-08-746-559A-5	Sequence 5, Appl
31	210.5	9.2	383	4	US-08-857-076-105	Sequence 105, Appl
32	203	8.9	1724	4	US-08-857-076-12	Sequence 12, Appl
33	196	8.6	366	4	US-08-857-076-103	Sequence 103, Appl
34	184.5	8.1	370	4	US-08-857-076-104	Sequence 104, Appl
35	147.5	6.4	381	4	US-08-857-076-106	Sequence 106, Appl
36	142	6.2	1940	2	US-08-644-271-30	Sequence 30, Appl
37	131.5	5.7	799	2	US-08-525-940-23	Sequence 23, Appl
38	131.5	5.7	799	2	US-08-976-838-23	Sequence 21, Appl
39	131.5	5.7	881	2	US-08-525-940-21	Sequence 21, Appl
40	131.5	5.7	881	2	US-08-976-838-21	Sequence 21, Appl
41	131.5	5.7	915	2	US-08-525-940-18	Sequence 18, Appl
42	131.5	5.7	915	2	US-08-976-838-18	Sequence 18, Appl
43	126	5.5	288	1	US-08-368-852-15	Sequence 15, Appl
44	124	5.4	288	2	US-08-525-940-15	Sequence 15, Appl
45	124	5.4	288	2	US-08-976-838-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-146-283-4
Sequence 4, Application US/09146283
Patent No. 5976546
GENERAL INFORMATION:
APPLICANT: Laus, Reiner
APPLICANT: Ruegg, Curtiss L.
TITLE OF INVENTION: Immunostimulatory Compositions
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Ave. Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/146,283
FILING DATE: 03-SEPT-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Judge, Linda R.
REGISTRATION NUMBER: 42,702
REFERENCE/DOCKET NUMBER: 7636-0010.21
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: homo sapiens
INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
US-09-146-283-4
Query Match 82.1%; Score 1878; DB 2; Length 782;
Best Local Similarity 83.0%; Pred. No. 5,1e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRWGLLLALLPFGAASVOTCTGDMKRLPASPETHLMDLRHLHYGCOVVOGNTL 60
DB 1 MELAALCRWGLLLALLPFGAASVOTCTGDMKRLPASPETHLMDLRHLHYGCOVVOGNTL 60
QY 61 ELTYLPTNASLSFLDIOGVGVYLLAHNOVQVPLQRLRIYRGQLFEDNALAVLNG 120
DB 61 ELTYLPTNASLSFLDIOGVGVYLLAHNOVQVPLQRLRIYRGQLFEDNALAVLNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYODTIIMKDFHKNQOLA 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYODTIIMKDFHKNQOLA 180
QY 181 LTLIDNRSRACHPCSPMKGSRWGESSEDCQSLRTVCAGGACARCKPLPTDCHEQC 240
DB 181 LTLIDNRSRACHPCSPMKGSRWGESSEDCQSLRTVCAGGACARCKPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFCASCTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFCASCTACP 300
QY 301 YNYLSTDVGSCTLVCPHLHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVVP 355
DB 301 YNYLSTDVGSCTLVCPHLHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVVP 355
QY 356 LRMQPG--PAHPYLSFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 356 LRMQPG--PAHPYLSFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
QY 361 IDFFACCKRIFGSLAPLPSPFQDPPASNT---APLQPEQLOVEFTEITGLYIISAWPD 417
DB 361 IDFFACCKRIFGSLAPLPSPFQDPPASNT---APLQPEQLOVEFTEITGLYIISAWPD 417
QY 406 --PDAHVAVNLRYEG 419
DB 406 --PDAHVAVNLRYEG 419
QY 418 SLPDLSVFQNLQVIRG 433
DB 418 SLPDLSVFQNLQVIRG 433

RESULT 2

US-08-579-823A-4
Sequence 4, Application US/08579823A
Patent No. 6080409
GENERAL INFORMATION:
APPLICANT: Laus, Reiner
APPLICANT: Ruegg, Curtis L.
APPLICANT: Wu, Hongyu
TITLE OF INVENTION: Immunostimulatory Composition and Method
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave. Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,823A
FILING DATE: 03-DEC-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Judge, Linda R.
REGISTRATION NUMBER: 42,702
REFERENCE/DOCKET NUMBER: 7636-0010
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

ORIGINAL SOURCE:
ORGANISM: homo sapiens
INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
US-08-579-823A-4

Query Match 82.1%; Score 1878; DB 3; Length 782;
Best Local Similarity 83.0%; Pred. No. 5,1e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRWGLLLALLPFGAASVOTCTGDMKRLPASPETHLMDLRHLHYGCOVVOGNTL 60
DB 1 MELAALCRWGLLLALLPFGAASVOTCTGDMKRLPASPETHLMDLRHLHYGCOVVOGNTL 60
QY 61 ELTYLPTNASLSFLDIOGVGVYLLAHNOVQVPLQRLRIYRGQLFEDNALAVLNG 120
DB 61 ELTYLPTNASLSFLDIOGVGVYLLAHNOVQVPLQRLRIYRGQLFEDNALAVLNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYODTIIMKDFHKNQOLA 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYODTIIMKDFHKNQOLA 180
QY 181 LTLIDNRSRACHPCSPMKGSRWGESSEDCQSLRTVCAGGACARCKPLPTDCHEQC 240
DB 181 LTLIDNRSRACHPCSPMKGSRWGESSEDCQSLRTVCAGGACARCKPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFCASCTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFCASCTACP 300
QY 301 YNYLSTDVGSCTLVCPHLHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVVP 355
DB 301 YNYLSTDVGSCTLVCPHLHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVVP 355
QY 356 LRMQPG--PAHPYLSFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 356 LRMQPG--PAHPYLSFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
QY 361 IDFFACCKRIFGSLAPLPSPFQDPPASNT---APLQPEQLOVEFTEITGLYIISAWPD 417
DB 361 IDFFACCKRIFGSLAPLPSPFQDPPASNT---APLQPEQLOVEFTEITGLYIISAWPD 417
QY 406 --PDAHVAVNLRYEG 419
DB 406 --PDAHVAVNLRYEG 419
QY 418 SLPDLSVFQNLQVIRG 433
DB 418 SLPDLSVFQNLQVIRG 433

RESULT 3

US-09-344-195-4
Sequence 4, Application US/09344195
Patent No. 6210662
GENERAL INFORMATION:
APPLICANT: Laus, Reiner
APPLICANT: Ruegg, Curtis L.
APPLICANT: Wu, Hongyu
TITLE OF INVENTION: Immunostimulatory Compositions
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave. Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/344,195
FILING DATE: 24-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/146,283
FILING DATE: 03-SEPT-1998
ATTORNEY/AGENT INFORMATION:

APPLICANT: Disis, Mary L.
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/414,417B
FILING DATE: 31-MAR-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sharkey, Richard G.
REGISTRATION NUMBER: 32,629
REFERENCE/DOCKET NUMBER: 920010.448C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-414-417B-68

Query Match 82.1%; Score 1878; DB 1; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELALCWMGLLLALPPGAASSTOVCSTGDMKRLRPASPTHLDMLRHLHYGCGVOGSL 60
DB 1 MELALCWMGLLLALPPGAASSTOVCSTGDMKRLRPASPTHLDMLRHLHYGCGVOGSL 60
QY 61 ELTYLPTNASLSFLDIOIEVOGYVLIANNOYROYPLQRLRIYRGTOLEFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOIEVOGYVLIANNOYROYPLQRLRIYRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELDLSLTETILKGGVLIQRNPQLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELDLSLTETILKGGVLIQRNPQLCYODTILMKDIFHKNNOLA 180
QY 181 LTLIDTNSRACHPCSPCKSGRCGSESEDCQSLTRTYCAGGACRCKGPLETDCCHQC 240
DB 181 LTLIDTNSRACHPCSPCKSGRCGSESEDCQSLTRTYCAGGACRCKGPLETDCCHQC 240
QY 241 AAGCTGPKHSDCLALHNHSGICELHCPALVTYNTDFEESPNREGRTFCASCVTACP 300
DB 241 AAGCTGPKHSDCLALHNHSGICELHCPALVTYNTDFEESPNREGRTFCASCVTACP 300
QY 301 YNYLSTDVGSCITVCPPLHNOEYTAEDGTQRCCKSPCAR---GTHSLPRPAVPVP 355
DB 301 YNYLSTDVGSCITVCPPLHNOEYTAEDGTQRCCKSPCAR---GTHSLPRPAVPVP 355
QY 356 LRMQCG--PAHPVLSFLRPMNDVSAFSLPLAPLSPTSVPL-----STVSGRGPD 405
DB 356 LRMQCG--PAHPVLSFLRPMNDVSAFSLPLAPLSPTSVPL-----STVSGRGPD 405
QY 406 --PDAHVAVNLSTRYEG 419
DB 406 --PDAHVAVNLSTRYEG 419
QY 418 SLPLDLSVFQNLQVING 433
DB 418 SLPLDLSVFQNLQVING 433

RESULT 6
US-08-484-438-8
Sequence 8, Application US/08484438
Patent No. 581098 5780031
GENERAL INFORMATION:
APPLICANT: Plowman, Gregory D.
APPLICANT: Culouscou, Jean-Michel
APPLICANT: Shoyab, Mohammed
APPLICANT: Siegal, Clay B.
APPLICANT: Hellstr m, Inggerd
APPLICANT: Hellstr m, Karl E.
TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penzie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-438-8

Query Match 82.1%; Score 1878; DB 2; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELALCWMGLLLALPPGAASSTOVCSTGDMKRLRPASPTHLDMLRHLHYGCGVOGSL 60
DB 1 MELALCWMGLLLALPPGAASSTOVCSTGDMKRLRPASPTHLDMLRHLHYGCGVOGSL 60
QY 61 ELTYLPTNASLSFLDIOIEVOGYVLIANNOYROYPLQRLRIYRGTOLEFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOIEVOGYVLIANNOYROYPLQRLRIYRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELDLSLTETILKGGVLIQRNPQLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELDLSLTETILKGGVLIQRNPQLCYODTILMKDIFHKNNOLA 180

```

OY 181 LTLIDPNSRACHPSPCKGRCGSESEDOQSLFTVCAGGCAKCGPLPTDCCHQC 240
Db 181 LTLIDPNSRACHPCSPCKGRCGSESEDOQSLFTVCAGGCAKCGPLPTDCCHQC 240
OY 241 AAGCTGPKRSDCLACLFHNHSGICELHCPALVTYNTDFFESMPNEGRYTGASCVTAC 300
Db 241 AAGCTGPKRSDCLACLFHNHSGICELHCPALVTYNTDFFESMPNEGRYTGASCVTAC 300
OY 301 YNYLSTDVSGCLVCPPLNNOEYTAEDGTORECKCKPCAR-----GTSLLPRAAVVP 355
Db 301 YNYLSTDVSGCLVCPPLNNOEYTAEDGTORECKCKPCARCYGLGMEHLEVRVYTSAN 360
OY 356 LEMQSG--PAHVLSFLRPSMDLVSAFVSLPLAPLSPTSVPI-----SPVSYGRGP 405
Db 361 IOEFAGCKKIGSGSLAFLESEFDGDPASNT---APLQPEQLOVFEETLEETGLYISAMP 417
OY 406 --PDAAVAVNLRRYEG 419
Db 418 SLPDLVSFQNLQVIRG 433

RESULT 7
US-08-486-348A-68
; Sequence 68, Application US/08486348A
; Patent No. 5846538
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disls, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharky, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010,448C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-486-348A-68

Query Match 82.1%; Score 1878; DB 2; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9,5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY 1 MELAALCFMGLLALLPPGAASVQCTGDKMLRLPASPEFHLDMLRLRYOGCOVQGNL 60
Db 1 MELAALCFMGLLALLPPGAASVQCTGDKMLRLPASPEFHLDMLRLRYOGCOVQGNL 60
OY 61 ELLTYPLTNASLSFLODIOEVQGYVLIANQVRYQVPLQRLRIYRGTOLEFDNYVALAVIDNG 120

```

```

Db      61 ELTYLPTNAAISLFQDDIQEVOGYVLLAHNOYRQPIQLRLRIYANGTDLFEDNNALAVLDNG 120
        |||||||
Qy      121 DPLNNTPVTCASPGGLREIQLNSLEILKGVLIORNPOLCYODTIIMKDIFHKNNOLA 180
        |||||||
Db      121 DPLNTPVTVCASPGGIREIQLNSTLEITLKGGVLIORNPOLCYODTIIMKDJFHKKNNOLA 180
        |||||||
Qy      181 LTLIDTNRSRACHCSPMKCKGRWGWSSSDCOSLTRVCAGGCACRCAPLPDCCHEQC 240
        |||||||
Db      181 LTLIDTNRSRACHCSPMKCKGRWGWSSSDCOSLTRVCAGGCACRCAPLPDCCHEQC 240
        |||||||
Qy      241 AAGCTGPRHSCLACLPHNHSGICELCPALVYNTDFESMNPREGRTTFGASCVTACP 300
        |||||||
Db      241 AAGCTGPRHSCCLACLPFNHSGICELCPALVYNTDFESMNPREGRTTFGASCVTACP 300
        |||||||
Qy      301 YNYISTDVGSCTLCPLHNOEVTAEDGTORCEKCSKPCAR----GTHSLPRAVPAYPV 355
        |||||
Db      301 YNYISTDVGSCTLCPLHNOEVTAEDGTORCEKCSKPCARVCYGICMEHLREVAVTSAN 360
        |||||
Qy      356 LRMPQG--PAHPVLSELRPSNDVSATYSILPALPLSTSVPI-----SVYSVGREDP 405
        ::::|
Db      361 IQEFAGCKRKIRFGSLAFPLESPFDGPASNT---APIQPLOLYETLEITGYLVISAMPD 417
        :|:|:|
Qy      406 --PDHAHVAVNLSTRYE 419
        |||
Db      418 SLPLDSVFQNLOVTRG 433
        |

RESULT      8
US-08-625-101-2
; Sequence 2, Application US/08625101
; Patent No. 5869445
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: COMPOUNDS FOR Eliciting OR ENHANCING IMMUNE
; TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION
; TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6500 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,101
; FILING DATE: 01-Apr-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-625-101-2

```

Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```

QY 1 MELAALCRNGLLALLPPCAASCTGCTDMKRLPASPETHLMDLRHLGYCCQVVGNTL 60
DB 1 MELAALCRNGLLALLPPCAASCTGCTDMKRLPASPETHLMDLRHLGYCCQVVGNTL 60
QY 61 ELTYLPTNLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTOLEFDNALAVLDNG 120
DB 61 ELTYLPTNLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTOLEFDNALAVLDNG 120
QY 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNOLA 180
DB 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNOLA 180
QY 181 LTLIDTNRBRACHPCSPMKGSRGWESSEDCOSLTRVCAGGACARCKPLPTDCCHQC 240
DB 181 LTLIDTNRBRACHPCSPMKGSRGWESSEDCOSLTRVCAGGACARCKPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTTFCASCTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTTFCASCTACP 300
QY 301 YNVLSTDVGSCTLVCPPLHNOEVTAEEDGTORCEKSKPCAR----GTHSLRPPAAYVP 355
DB 301 YNVLSTDVGSCTLVCPPLHNOEVTAEEDGTORCEKSKPCARVCYGLGMEHLREVRATVSAN 360
QY 356 LKMQRG--PAHVLSTLRSMPLVSFYLPLAPLSPTVPI-----SPVSVGRGD 405
DB 356 LKMQRG--PAHVLSTLRSMPLVSFYLPLAPLSPTVPI-----SPVSVGRGD 405
QY 406 --PDAHVAVNLRSRYEG 419
DB 406 --PDAHVAVNLRSRYEG 419
QY 418 SLPDLSVFONLOYIRG 433
DB 418 SLPDLSVFONLOYIRG 433

```

RESULT 9

US-08-468-545B-68

Sequence 68, Application US/08468545B

Patent No. 5876712

GENERAL INFORMATION:

APPLICANT: Cheever, Martin A.

APPLICANT: Disis, Mary L.

TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN

TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE

TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed and Berry LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: US

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,545B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Sharkey, Richard G.

REGISTRATION NUMBER: 32,629

REFERENCE/DOCKET NUMBER: 920010.448C5

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:

LENGTH: 1255 amino acids

TYPE: amino acid
TOPOLOGY: linear
US-08-468-545B-68

Query Match 82.1%; Score 1878; DB 2; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```

QY 1 MELAALCRNGLLALLPPCAASCTGCTDMKRLPASPETHLMDLRHLGYCCQVVGNTL 60
DB 1 MELAALCRNGLLALLPPCAASCTGCTDMKRLPASPETHLMDLRHLGYCCQVVGNTL 60
QY 61 ELTYLPTNLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTOLEFDNALAVLDNG 120
DB 61 ELTYLPTNLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTOLEFDNALAVLDNG 120
QY 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNOLA 180
DB 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNOLA 180
QY 181 LTLIDTNRBRACHPCSPMKGSRGWESSEDCOSLTRVCAGGACARCKPLPTDCCHQC 240
DB 181 LTLIDTNRBRACHPCSPMKGSRGWESSEDCOSLTRVCAGGACARCKPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTTFCASCTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTTFCASCTACP 300
QY 301 YNVLSTDVGSCTLVCPPLHNOEVTAEEDGTORCEKSKPCAR----GTHSLRPPAAYVP 355
DB 301 YNVLSTDVGSCTLVCPPLHNOEVTAEEDGTORCEKSKPCARVCYGLGMEHLREVRATVSAN 360
QY 356 LKMQRG--PAHVLSTLRSMPLVSFYLPLAPLSPTVPI-----SPVSVGRGD 405
DB 356 LKMQRG--PAHVLSTLRSMPLVSFYLPLAPLSPTVPI-----SPVSVGRGD 405
QY 406 --PDAHVAVNLRSRYEG 419
DB 406 --PDAHVAVNLRSRYEG 419
QY 418 SLPDLSVFONLOYIRG 433
DB 418 SLPDLSVFONLOYIRG 433

```

RESULT 10

US-08-356-786-2

Sequence 2, Application US/08356786

Patent No. 5877305

GENERAL INFORMATION:

APPLICANT: Huston, James S.

APPLICANT: Oppermann, Hermann

APPLICANT: Houston, L. L.

APPLICANT: Ring, David B.

TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer

TITLE OF INVENTION: Marker

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault

STREET: Exchange Place, 53 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/356,786

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/831,967

FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:
 NAME: Pitcher, Edmund R.
 REGISTRATION NUMBER: 27,829
 REFERENCE/DOCKET NUMBER: CRP-053
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 248-7000
 TELEFAX: (617) 248-7100
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1255 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-356-786-2

Query Match 82.1%; Score 1878; DB 2; Length 1255;
 Best Local Similarity 83.0%; Pred. No. 9.5e-156;
 Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METALACRMGILLALLPPGASTQVCTGDMKRLPASPEHLDMLRHYGCCQVVGNTL 60
 DB 1 METALACRMGILLALLPPGASTQVCTGDMKRLPASPEHLDMLRHYGCCQVVGNTL 60
 QY 61 ELTYLPTNASLFLDIDIEVQGYVLIANQVQVPLQRLRIYRGQLFEDNVALAVLNG 120
 DB 61 ELTYLPTNASLFLDIDIEVQGYVLIANQVQVPLQRLRIYRGQLFEDNVALAVLNG 120
 QY 121 DPLNNTTPTVGTASPGGLRELDRLSLTEILKGVLIIQRNPOLCYQDTILMKDIFHKNNOLA 180
 DB 121 DPLNNTTPTVGTASPGGLRELDRLSLTEILKGVLIIQRNPOLCYQDTILMKDIFHKNNOLA 180
 QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRGCPLEPTDCCHQC 240
 DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRGCPLEPTDCCHQC 240
 QY 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
 DB 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 QY 366 LRMQPG--PAHPVLSFLRPMNDVSAFYSPLAPLSPTSVPL-----SPVSYGRGPD 405
 DB 361 IQEFAGCKRKIFGSLAFLEPSFDGDPASNT---APLQPOLQVFTLEITGYLISAMPD 417
 QY 406 --PDAHVAVNLSRYEG 419
 DB 418 SLPLDLSVFQNLQVIRG 433

RESULT 11
 US-08-466-680B-68
 Sequence 68, Application US/08466680B
 Patent No. 6075122
 GENERAL INFORMATION:
 APPLICANT: Cheever, Martin A.
 APPLICANT: Disis, Mary L.
 TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
 TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
 NUMBER OF SEQUENCES: 69
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: US
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/466,680B
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Sharkey, Richard G.
 REGISTRATION NUMBER: 32,629
 REFERENCE/DOCKET NUMBER: 920010.448C4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 68:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1255 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-466-680B-68

Query Match 82.1%; Score 1878; DB 3; Length 1255;
 Best Local Similarity 83.0%; Pred. No. 9.5e-156;
 Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METALACRMGILLALLPPGASTQVCTGDMKRLPASPEHLDMLRHYGCCQVVGNTL 60
 DB 1 METALACRMGILLALLPPGASTQVCTGDMKRLPASPEHLDMLRHYGCCQVVGNTL 60
 QY 61 ELTYLPTNASLFLDIDIEVQGYVLIANQVQVPLQRLRIYRGQLFEDNVALAVLNG 120
 DB 61 ELTYLPTNASLFLDIDIEVQGYVLIANQVQVPLQRLRIYRGQLFEDNVALAVLNG 120
 QY 121 DPLNNTTPTVGTASPGGLRELDRLSLTEILKGVLIIQRNPOLCYQDTILMKDIFHKNNOLA 180
 DB 121 DPLNNTTPTVGTASPGGLRELDRLSLTEILKGVLIIQRNPOLCYQDTILMKDIFHKNNOLA 180
 QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRGCPLEPTDCCHQC 240
 DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRGCPLEPTDCCHQC 240
 QY 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
 DB 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 QY 366 LRMQPG--PAHPVLSFLRPMNDVSAFYSPLAPLSPTSVPL-----SPVSYGRGPD 405
 DB 361 IQEFAGCKRKIFGSLAFLEPSFDGDPASNT---APLQPOLQVFTLEITGYLISAMPD 417
 QY 406 --PDAHVAVNLSRYEG 419
 DB 418 SLPLDLSVFQNLQVIRG 433

RESULT 12
 US-08-422-108-1
 Sequence 1, Application US/08422108
 Patent No. 6015567
 GENERAL INFORMATION:
 APPLICANT: Huddiak, Robert M.
 APPLICANT: Shepard, H. Michael
 APPLICANT: Ullrich, Axel
 TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco

STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/422,108
FILING DATE: 14-Apr-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/355460
FILING DATE: 13-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/048346
FILING DATE: 15-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/354319
FILING DATE: 19-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 554C2D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-422-108-1

Query Match 77.4%; Score 1769; DB 3; Length 624;
Best Local Similarity 82.2%; Pred. No. 1.3e-146;
Matches 341; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
QY 22 STQVCTGDMKRLRPASPEHMLRHLVQGVQVQVGNLELYLPTNASLFLQDIOEQV 81
DB 1 LLAALCPASRALEEKVKVCGTGNKLTQGTFFEDHFLQRMNCEVVLGNLEITYVQRN 73
QY 82 GYVLAHNOVROVPLQRLRIYVGTQLEFENYALAVLDNDPPLNNTPTVGSAGGIREQ 141
DB 61 GYVLAHNOVROVPLQRLRIYVGTQLEFENYALAVLDNDPPLNNTPTVGSAGGIREQ 120
QY 142 LRSLEILKGVLIQNPOLCYODTILMKDIFHKNNQALATLIDNRSRACHPCSPMCKG 201
DB 121 LRSLEILKGVLIQNPOLCYODTILMKDIFHKNNQALATLIDNRSRACHPCSPMCKG 180
QY 202 SRQWSESSDCCSLRTYVAGGACARCKGPLPTDCCHQCCAGCTGPKHSDCLACHFNHS 261
DB 181 SRQWSESSDCCSLRTYVAGGACARCKGPLPTDCCHQCCAGCTGPKHSDCLACHFNHS 240
QY 262 GICEHLCALVLYNNDTPESMPNPGRTYFGASCYACRYNLSLDVSGCTLVCPHLANE 321
DB 241 GICEHLCALVLYNNDTPESMPNPGRTYFGASCYACRYNLSLDVSGCTLVCPHLANE 300
QY 322 VAAEDGTORCEKSPCARVCGIGMEHLREYRAVTSANIQEPAGCKITFGSLAFPLESE 374
DB 301 VAAEDGTORCEKSPCARVCGIGMEHLREYRAVTSANIQEPAGCKITFGSLAFPLESE 360
QY 375 DLVSATYSLPLAPLSPTSVPI-----SPYVSGRGP--PDAAVAVMLSRYEG 419
DB 361 DGDPRASMT---APLQPEQVQFETLEETREYLYISAMPDLSLDFVFOQLQYIRG 412

RESULT 13
US-08-336-708a-9
Sequence 9, Application US/08336708A

Patent No. 5521295
GENERAL INFORMATION:
APPLICANT: Pacificl, Robert E.
APPLICANT: Thomason, Arlen R.
TITLE OF INVENTION: Hybrid Receptor Molecules
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,708A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy
REFERENCE/DOCKET NUMBER: A-241A
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 644 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-708a-9

Query Match 34.7%; Score 793; DB 1; Length 644;
Best Local Similarity 45.3%; Pred. No. 2.9e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;
QY 11 LLLALLPGAA--STQVCTGDMKRLRPASPEHMLRHLVQGVQVQVGNLELYLPTN 68
DB 14 LLAALCPASRALEEKVKVCGTGNKLTQGTFFEDHFLQRMNCEVVLGNLEITYVQRN 73
QY 69 ASLSFLQDIOEQVGYVLAHNOVROVPLQRLRIYVGTQLEFENYALAVLDNDPPLNNTPT 128
DB 74 YDLSFLKTIQEVAGYVLAHNOVROVPLQRLRIYVGTQLEFENYALAVLDNDPPLNNTPT 126
QY 129 VTGASPGGLREQLRSLTEILKGVLIQNPOLCYODTILMKDIFHKNNQALATLIDNRS 188
DB 127 ---AKTIGLKEPLMNLQELILGAVRFSSNPALCNVESIQWRDIYSSDFLSMSPDFQNH 183
QY 189 SRACHPCSPMCKGSRQWSESSDCCSLRTYVAGGCA-RCKGPLPTDCCHQCCAGCTGP 247
DB 184 LSSCCKCPSPCNGSCWGAEECCOKLTKIICAOCCSGRCRCKSPSDCHNCCAGCTGP 243
QY 248 KHSDDLACHFNHSGICEHLCALVLYNNDTPESMPNPGRTYFGASCYACRYNLSLDVSG 307
DB 244 RESDCLVCKEFDEATCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCVKCPRYVYVD 303
QY 308 VSGCTLVCPHLNQEVTAAEDGTORCEKSPCAR 340
DB 304 HGSYVRACGADSYEM-EDGVRKCKCKGSPCKR 335

RESULT 14
US-08-484-438-7
Sequence 7, Application US/08484438
Patent No. 581098
Patent No. 581098 5780031
GENERAL INFORMATION:
APPLICANT: Plouman, Gregory D.
APPLICANT: Culouscou, Jean-Michel

APPLICANT: Shoyab, Mohammed
APPLICANT: Siegal, Clay B.
APPLICANT: Helistr m, Ingerd
APPLICANT: Helistr m, Karl E.
TITLE OF INVENTION: HERA HUMAN RECEPTOR TYROSINE KINASE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-438-7

Query Match 34.7%; Score 793; DB 2; Length 1210;
Best Local Similarity 45.3%; Pred. No. 6,7e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

11 LLLALLPAGA--STOVCTGTDMLRLPASPEHLDMLRLHYOGCOVGNLELTPTN 68
14 LLAALCPASRALEKKVCGTSNKLTOLGTFEDHFLSLQRMNCEVYVGNLEITTYQNR 73
69 ASLSFLQDIOEVGYVLAHNOVROYPLRLRIVGTQLFEDNYALAVLDNGPLNNTTP 128
74 YDLSFLKTIOEVGYVLAHNOVROYPLRLRIVGTQLFEDNYALAVLDNGPLNNTTP 126
129 VTGASPGGLRELOLSITELIKGVLIOIRNPOLCYODTILMKDIFKNNQALATLIDTNR 188
127 ---ANKTGLKEIPKRLDELILHGAVERFSNNPALCNVESTIOMRDIVSSDFLSNKMDFQNH 183
189 SRACHPCSPMKSGRCWGSSESDQSLRTTYCAGGCA-RCKGPLPTDCCHEQCAACTGP 247
184 LSGCQKCDPSCPNCGSGWAGEENCKLTKIICAGQCSGRCGRKSPDCCCHNCAACTGP 243
248 KHSDDLACHFNHSGICELHCPALVTYNTDTFESMNPBGRITFFGASCYATACPYNTLSD 307
244 RESDCLVCRKFRDEATCKDTCPLMLYNPTTYOMDVNPGKYSFGATCVKCKPRNRYVVD 303

308 VGSCTLVCPHLNHOEYTAEDGTORCEKSKPCAR 340
304 HGSVCVRACGADSYEM-BEDGVRCRCKCEGPCRK 335
RESULT 15
US-08-475-035-4
Sequence 4, Application US/08475035
Patent No. 5985353
GENERAL INFORMATION:
APPLICANT: KING, C. R.
APPLICANT: KRAUS, MATTHIAS H.
APPLICANT: AARONSON, STUART A.
TITLE OF INVENTION: HUMAN GENE RELATED TO BUT DISTINCT FROM
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: Suite 1200, 127 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,035
FILING DATE: 7 Jun 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perryman, David G.
REGISTRATION NUMBER: 33,438
REFERENCE/DOCKET NUMBER: 1414.656
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-035-4

Query Match 34.7%; Score 793; DB 2; Length 1210;
Best Local Similarity 45.3%; Pred. No. 6,7e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

11 LLLALLPAGA--STOVCTGTDMLRLPASPEHLDMLRLHYOGCOVGNLELTPTN 68
14 LLAALCPASRALEKKVCGTSNKLTOLGTFEDHFLSLQRMNCEVYVGNLEITTYQNR 73
69 ASLSFLQDIOEVGYVLAHNOVROYPLRLRIVGTQLFEDNYALAVLDNGPLNNTTP 128
74 YDLSFLKTIOEVGYVLAHNOVROYPLRLRIVGTQLFEDNYALAVLDNGPLNNTTP 126
129 VTGASPGGLRELOLSITELIKGVLIOIRNPOLCYODTILMKDIFKNNQALATLIDTNR 188
127 ---ANKTGLKEIPKRLDELILHGAVERFSNNPALCNVESTIOMRDIVSSDFLSNKMDFQNH 183
189 SRACHPCSPMKSGRCWGSSESDQSLRTTYCAGGCA-RCKGPLPTDCCHEQCAACTGP 247
184 LSGCQKCDPSCPNCGSGWAGEENCKLTKIICAGQCSGRCGRKSPDCCCHNCAACTGP 243
248 KHSDDLACHFNHSGICELHCPALVTYNTDTFESMNPBGRITFFGASCYATACPYNTLSD 307
244 RESDCLVCRKFRDEATCKDTCPLMLYNPTTYOMDVNPGKYSFGATCVKCKPRNRYVVD 303

OY 308 VGSCTLVCPHNOEYTAEDGTORCEKCSKPCAR 340
||| | : | : | : | : | :
DB 304 HGSCTVRACGADSYEM-EBDGVRRCKKCEGPCRK 335

Search completed: April 11, 2002, 09:29:38
Job time: 75 sec